

Das IAB-Betriebspanel - von der Stichprobe über die Erhebung bis zur Hochrechnung

Fischer, Gabriele; Janik, Florian; Müller, Dana; Schmucker, Alexandra

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
SSG Sozialwissenschaften, USB Köln

Empfohlene Zitierung / Suggested Citation:

Fischer, G., Janik, F., Müller, D., & Schmucker, A. (2008). *Das IAB-Betriebspanel - von der Stichprobe über die Erhebung bis zur Hochrechnung*. (FDZ Methodenreport, 1/2008). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit (IAB) Forschungsdatenzentrum (FDZ). <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-364060>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. By using this particular document, you accept the above-stated conditions of use.

Das IAB-Betriebspanel – von der Stichprobe über die Erhebung bis zur Hochrechnung

Gabriele Fischer (TNS Infratest Sozialforschung)

Florian Janik (IAB Forschungsbereich Betriebe und Beschäftigung)

Dana Müller (Forschungsdatenzentrum der BA im IAB)

Alexandra Schmucker (Forschungsdatenzentrum der BA im IAB)

Änderungen zur Erstfassung

Der FDZ-Methodenreport 01/2008 wurde am 22.04.2008 aktualisiert. Die Neufassung beinhaltet redaktionelle Änderungen in den Kapiteln 2.2 und 4.2.

Danksagung

Für Kritik, Anregungen und Ergänzungen danken wir Lutz Bellmann, Harald Bielski, Udo Brixy, Micha Andreas Crimmann, Peter Ellguth, Nicole Gürtzgen, Tanja Hethey, Martina Huber, Peter Jacobebbinghaus, Susanne Kohaut, Johannes Ludsteck und Dieter Urban. Für den Inhalt bleiben die Autoren allein verantwortlich.

Inhaltsverzeichnis

1. Einleitung.....	4
2. Stichprobendesign.....	6
2.1 Die Betriebsnummer.....	7
2.2 Die Stichprobenziehung.....	8
2.3 Die Teilstichproben des IAB-Betriebspanels.....	9
3. Erhebungsdesign, Feldarbeit und Editing	12
3.1 Fragebogenkonzeption und Pretest.....	12
3.2 Erhebungsmethode.....	14
3.3 Feldphase und Identifikation der richtigen Betriebseinheit.....	14
3.4 Ausfälle bei Interviews und Fragen	16
3.5 Datenprüfung (Editing).....	17
4. Auswertungen im Querschnitt.....	18
4.1 Definition der Querschnittsfälle	19
4.2 Querschnittshochrechnung.....	19
4.3 Hinweise zur Auswertungen im Querschnitt.....	21
5. Auswertungen im Längsschnitt	23
5.1 Definition von Längsschnittsfällen	24
5.2 Längsschnittshochrechnung.....	25
5.3 Hinweise zur Auswertung im Längsschnitt.....	27
5.4 Das Wellenkennzeichen: Definition eigener Längsschnitte.....	28
6. Datenzugang.....	31
Literatur.....	33
Anhang.....	34

1. Einleitung

Das IAB-Betriebspanel ist eine für Deutschland einzigartige, jährliche Betriebsbefragung, da sie bundesweit alle Branchen und Betriebsgrößen repräsentiert und auch im Längsschnitt ausgewertet werden kann. Das Design für das IAB-Betriebspanel wurde zu Beginn der 90er Jahre entwickelt und umfangreich getestet. Dabei wurde auch auf parallel laufende Entwicklungsaktivitäten im Rahmen des Hannoveraner Firmenpanels zurückgegriffen, die damals im Auftrag der Forschungsstelle Firmenpanel an der Universität Hannover (Gerlach et al.: 1998) und des IAW Tübingen durchgeführt wurden.

Die Befragung begann 1993 in Westdeutschland mit dem Ziel, ein repräsentatives Informationssystem zur kontinuierlichen Analyse der Arbeitskräftenachfrage aufzubauen. Seit 1996 wird die Erhebung auch in Ostdeutschland und somit bundesweit durchgeführt. Das IAB-Betriebspanel ist als Längsschnitterhebung konzipiert, d.h. jedes Jahr wird ein Großteil derselben Betriebe wieder befragt. Folglich können sowohl Entwicklungen im Zeitverlauf durch den Vergleich von Querschnittsdaten zu unterschiedlichen Zeitpunkten analysiert, als auch betriebsindividuelle Längsschnittuntersuchungen vorgenommen werden.

Mittlerweile werden knapp 16.000 Betriebe zu einer Vielzahl beschäftigungspolitischer Themen befragt, u. a. zur Beschäftigungsentwicklung, Geschäftspolitik und Geschäftsentwicklung, zu Investitionsaktivitäten, Innovationen im Betrieb, öffentlicher Förderung, Personalstruktur, Berufsausbildung und Ausbildungsstellen, Personaleinstellungen und -abgängen, Personalsuche, Löhnen und Gehältern, betrieblichen Arbeitszeiten, Fort- und Weiterbildung sowie zu allgemeinen Angaben zum Betrieb. Hinzu kommen jährlich wechselnde Themenschwerpunkte. Mit Ausnahme von Hamburg beteiligen sich aktuell alle Bundesländer mit regionalen Aufstockungsstichproben am IAB-Betriebspanel. Dies ermöglicht zum einen Auswertungen auf der Bundeslandebene, zum anderen ergibt sich daraus in der Summe ein Stichprobenumfang, der auf Bundesebene die Auswertungsmöglichkeiten deutlich erhöht.

Das IAB-Betriebspanel zeichnet sich durch eine hohe Datenqualität aus, die mit der hochwertigen Stichprobe, der hohen Ausschöpfung, sowie mit dem aufwändigen Prozess von Datenprüfung und Fehlerkorrektur erreicht wird. Die Befragung wird von TNS Infratest Sozialforschung GmbH im Auftrag des IAB durchgeführt. Eine allgemeine Einführung zum IAB-Betriebspanel findet sich bei Bellmann (2002) oder in englischer Sprache bei Kölling (2000).

Das IAB-Betriebspanel basiert auf einem komplexen Studiendesign, das auch die Nutzer und Nutzerinnen des Datensatzes vor Herausforderungen stellt. Mit diesem Papier wird ein Überblick über die methodischen Besonderheiten des IAB-Betriebspanels gegeben. Im Einzelnen wird detaillierter auf das Stichproben- und Erhebungsdesign, die Hochrechnung und den Da-

tenzugang im Forschungsdatenzentrum (FDZ) der Bundesagentur für Arbeit (BA) am Institut für Arbeitsmarkt- und Berufsforschung (IAB) eingegangen. Die wichtigsten Punkte der einzelnen Kapitel werden jeweils zu Beginn vorgestellt. Den Nutzerinnen und Nutzern des IAB-Betriebspanels soll es zum einen als Sammlung von methodischen Aspekten des IAB-Betriebspanel dienen und zum anderen insbesondere für Erstnutzer und Erstnurtzerinnen den Einstieg in die Datennutzung erleichtern.

Dieses Papier richtet sich ebenso an die Nutzerinnen und Nutzer der Linked-Employer-Employee Daten des IAB (LIAB¹), für die das IAB-Betriebspanel einen wichtigen Bestandteil darstellt.

Der FDZ Datenreport 6/2006 (Städele/Müller 2006) enthält die Darstellung der einzelnen Variablen und deren Auszählungen für die Jahre 1993 bis 2005. Die Auszählungen aller Variablen der Welle 2006 befinden sich in einem zusätzlichen Dokument unter: http://doku.iab.de/fdz/iabb/Fallzahlen_IAB-Betriebspanel2006.zip. Einen Überblick über alle Variablen im Zeitverlauf bietet die Variablenliste 1993 bis 2006 (http://doku.iab.de/fdz/iabb/Variablenliste1993_2006.xls).

¹ Dieser Datensatz umfasst zum einen die Daten des IAB-Betriebspanels und zum anderen die prozessproduzierten Personendaten der Bundesagentur für Arbeit (BA). Dadurch ist eine gleichzeitige Analyse der Angebots- und Nachfrageseite des Arbeitsmarktes möglich (Alda/Bender/Gartner 2005).

2. Stichprobendesign

Das Wichtigste in Kürze

Die Grundgesamtheit des IAB-Betriebspanels umfasst alle Betriebe mit mindestens einem sozialversicherungspflichtig Beschäftigten zum 30.06. des Vorjahres.

Grundlage für die Stichprobenziehung ist die Betriebsdatei der Bundesagentur für Arbeit. Dies ist die einzige Datenquelle in Deutschland, die Betriebe aller Branchen und Betriebsgrößen umfasst. Aus der Betriebsdatei werden für die Stichprobe des IAB-Betriebspanels Betriebsnummern gezogen. Diese stellen grundsätzlich die Befragungseinheiten des IAB-Betriebspanels dar. Jedoch steckt nicht hinter jeder Betriebsnummer eine sinnvoll zu befragende Einheit.

Das IAB-Betriebspanel basiert auf einer nach Betriebsgröße, Branche und Bundesland disproportional geschichteten Stichprobe. Deskriptiv müssen gewichtete Daten verwendet werden, um repräsentative Aussagen über die Grundgesamtheit treffen zu können.

Bei Branchenanalysen ist darauf zu achten, dass durch die Veränderung der Wirtschaftszweigsystematik (v.a. von WS73 auf WZ93) die Vergleichbarkeit über die Zeit nur eingeschränkt gegeben ist. Vergleichende Branchenanalysen sind von 1993-1999 und ab 2000 bis zum aktuellen Rand möglich.

Die Gründung von Betrieben wird im IAB-Betriebspanel durch die Ziehung von „neuen“ Betriebsnummern abgebildet. Dabei handelt es sich jedoch nicht zwangsläufig um neu gegründete Betriebe.

Die Grundgesamtheit des IAB-Betriebspanels umfasst alle Betriebe mit mindestens einem sozialversicherungspflichtig Beschäftigten zum 30.06. des Vorjahres.² Grundlage für die Stichprobenziehung ist die Betriebsdatei der BA, die vierteljährlich erstellt wird und rund zwei Millionen Betriebe umfasst. Die Betriebsdatei enthält alle Betriebe, die im Rahmen ihrer Meldepflicht die Beschäftigten des Betriebes an die Sozialversicherungsträger melden. Für diese Meldungen erhalten die Betriebe von der jeweils für den Betrieb zuständigen Agentur für Arbeit, ab 2007 von dem zentralen Betriebsnummernservice der BA, eine Betriebsnummer. In der Be-

² Seit 2004 werden keine privaten Haushalte (vorher ab einer Zahl von mindestens fünf Beschäftigten zum 30.06. des Vorjahres) und exterritorialen Einheiten (z. B. Botschaften) mehr befragt. Beide Arten von Einheiten sind nicht weit verbreitet. Hinzu kommt, dass die Befragung nicht für die Untersuchung privater Haushalte konzipiert wurde. Exterritoriale Einheiten sind nur in Ausnahmefällen bereit gewesen, sich an der Befragung zu beteiligen. Daher enthält die Grundgesamtheit seit 2004 ca. 30.000 Betriebe weniger, inklusive 1.000 Betriebsnummern, für die die Betriebsdatei keine Angaben enthält.

triebsdatei der BA werden diese Betriebsnummern zentral zusammengeführt. Für die Stichprobenziehung und Hochrechnung ist die Betriebsnummer die relevante Einheit.

2.1 Die Betriebsnummer

Eine Betriebsnummer ist zunächst eine für den Verwaltungsprozess notwendige Kenngröße. Jeder Betrieb bekommt im Rahmen des Meldeverfahrens der Sozialversicherungen eine Betriebsnummer zugewiesen. Unter „Betrieb“ wird in diesem Sinne eine regional und wirtschaftlich abgegrenzte Einheit verstanden, in der sozialversicherungspflichtig Beschäftigte tätig sind. Für die Vergabe von Betriebsnummern gelten bestimmte Regeln, die in den einzelnen Arbeitsagenturen Anwendung finden. Neue Betriebsnummern werden in der Regel vergeben, wenn:

- ein Betrieb neu gegründet wird,
- ein Wechsel des Betriebsinhabers³ stattfindet,
- die wirtschaftliche Zuordnung des Betriebs sich ändert oder
- ein Antrag vom Arbeitgeber z. B. auf Zusammenlegung von Filialen vorliegt.

Bei der Verlegung eines Betriebs in eine andere Gemeinde oder einen anderen Agenturbezirk wird keine neue Betriebsnummer zugeteilt, außer es liegt ein Antrag des Arbeitgebers oder der Krankenkasse vor. Betriebe, die eine Betriebsnummer haben, in denen jedoch vorübergehend keine sozialversicherungspflichtig Beschäftigten tätig sind, werden als „ruhend“ bezeichnet. Die alte Betriebsnummer kann wieder aufgenommen werden, wenn der Arbeitgeber und die Krankenkasse damit einverstanden sind. Wenn der Arbeitgeber einen neuen Betrieb eröffnet, kann er unter Umständen auf seine alte Betriebsnummer zurückgreifen.

In den meisten Fällen stellt die Betriebsnummer eine geeignete Grundlage für die Identifikation eines Betriebs als Befragungseinheit dar. Allerdings ist das nicht immer der Fall. Die oben genannten Kriterien für die Vergabe von Betriebsnummern implizieren beispielsweise, dass ein Unternehmen über mehrere Betriebsnummern verfügen kann, wenn die Unternehmenseinheiten in verschiedenen Agenturbezirken liegen oder unterschiedliche, voneinander inhaltlich getrennte wirtschaftliche Einheiten darstellen. Es gibt auch den umgekehrten Fall: Mehrere Niederlassungen desselben Unternehmens in einer Gemeinde werden in der Betriebsnummernsystematik zu einem Betrieb zusammengefasst, z. B. Hauptgeschäftsstelle und mehrere Filialen, wenn die gleiche Wirtschaftsklasse und das Einverständnis des Arbeitgebers sowie der Krankenkasse vorliegen.

³ Die Übergabe des Betriebs an Familienangehörige ist davon nicht betroffen, solange die Krankenkasse keine neue Betriebsnummer fordert.

Inwieweit es sich bei der gezogenen Betriebsnummer um eine sinnvoll zu befragende Einheit handelt, also um eine Einheit, für die die erwünschten Fragebogenangaben im Betrieb vorliegen, wird letztendlich von dem Interviewer/der Interviewerin vor Ort geklärt. Das Studiendesign des IAB-Betriebspanels lässt bei der Identifikation der zu befragenden Betriebseinheit gewisse Abweichungen zu. Es muss allerdings immer ein Bezug zur gezogenen Betriebsnummer vorhanden sein. Weitere Ausführungen zur Betriebsnummer und zur Identifikation der zu befragenden Betriebe sind im Kapitel 3.3 zu finden.

2.2 Die Stichprobenziehung

Die Stichprobenziehung des IAB-Betriebspanels erfolgt aus der Betriebsdatei⁴ zum 30. Juni des Vorjahres, weil die Informationen zu den Meldungen zur Sozialversicherung erst ca. sechs Monate nach dem jeweiligen Stichtag vorliegen. Jedoch liegt nicht für alle Betriebe die genaue Zahl der sozialversicherungspflichtig Beschäftigten zum Stichtag vor. Für ca. 5% der sozialversicherungspflichtig Beschäftigten fehlen aufgrund unterschiedlicher Meldezeitpunkte die Informationen zum Stichtag (Bundesagentur für Arbeit 2005). Für diese Fälle werden Informationen auf Basis der letzten vorliegenden Meldung fortgeschrieben.

Die Stichprobe ist nach Betriebsgröße⁵, Branche und Bundesland geschichtet. Es handelt sich dabei um eine mehrfach disproportionale Stichprobe. Die Sollbesetzung der Zellen der Schichtungsmatrix wird zum einen durch den Umfang der Aufstockungsstichproben (eine Übersicht zu den Aufstockungen in den einzelnen Bundesländern findet sich im Anhang in Tabelle E) und zum anderen nach dem „Prinzip der optimalen Schichtung“ (Buttler/Fickel 2002, S. 147f) über die Betriebsgrößenklassen beschäftigtenproportional bestimmt. Im Ergebnis sind große Betriebe, kleine Bundesländer und kleine Branchen sowie das Verarbeitende Gewerbe in Ostdeutschland⁶ überrepräsentiert (siehe Tabelle A im Anhang). Diese disproportionale Schichtung hat Konsequenzen für Auswertungen, die in den Kapiteln 4 und 5 beschrieben werden.

Die Schichtungsmatrix hat sich im Laufe der Zeit für die Branchen aufgrund von Veränderungen in der Wirtschaftszweig-Systematik (von WS73⁷ auf WZ93⁸) und modifizierter inhaltlicher Zielsetzungen gewandelt.

⁴ Weiterführende Informationen zur Betriebsdatei sind bei Brixy/Fritsch (2002) nachzulesen.

⁵ Im IAB-Betriebspanel wird für Stichprobe und Gewichtung unter dem Begriff Betriebsgröße die Anzahl der sozialversicherungspflichtig Beschäftigten im Betrieb verstanden.

⁶ Diese Aufstockungsstichprobe wird vom IW Halle finanziert.

⁷ Die WS73 ist eine eigens von der Bundesanstalt für Arbeit erstellte Wirtschaftszweigklassifikation (Bundesanstalt für Arbeit 1973), die sich von der Wirtschaftszweigklassifikation des Statistischen Bundesamtes (WZ79) unterscheidet (Bender et al. 1996: 16). Seit Einführung der WZ93 werden von der Bundesagentur für Arbeit und dem Statistischen Bundesamt einheitliche Klassifikationen verwendet.

Zu Beginn der Erhebung im Jahr 1993 umfasste die Schichtungsmatrix 16 Wirtschaftszweige, ab 2000 waren es 20 Branchen und seit 2004 sind es 17 Branchen (siehe Tabellen B bis D im Anhang). Diese Umstellungen sind bei branchenspezifischen Zeitreihenbetrachtungen zu berücksichtigen. Vor allem mit dem Wechsel von WS73 auf WZ93 im Jahr 2000 ergaben sich gravierende Veränderungen bei der Zuordnung der Betriebe, so dass Vergleiche über den gesamten Erhebungszeitraum nur eingeschränkt möglich sind. Auch über eine größere Klassifikation kann keine Vergleichbarkeit erreicht werden.

Bis 2006 wurde für jedes Bundesland und für West- und Ostberlin eine getrennte Stichprobe gezogen. Die genaue räumliche Zuordnung der Betriebsnummern nach West- und Ostberlin ist in der Betriebsdatei allerdings nicht mehr möglich, so dass seit 2007 Berlin als Ganzes zugrunde gelegt wird. Dies hat Auswirkungen auf die Definition von West- und Ostdeutschland. Bis einschließlich 2006 wurde Ostdeutschland mit Ostberlin und Westdeutschland mit Westberlin zusammengefasst, seit 2007 umfasst Westdeutschland alle westdeutschen Bundesländer ohne Berlin, während Ostdeutschland das gesamte Berlin zugeordnet wurde.⁹ Dies gilt für die Stichprobenziehung und damit auch analog für die Hochrechnung.

Für die Ziehung der Stichprobe werden die Sollzahlen für die Nettostichprobe in Abhängigkeit der finanzierten Interviews auf Bundeslandebene sowie für das Verarbeitende Gewerbe in Ostdeutschland ermittelt. Für diese 17 Teilstichproben (seit 2003) wird für jede Zelle der Schichtungsmatrix (17 Branchen und 10 Betriebsgrößenklassen) nach dem „Prinzip der optimalen Schichtung“ eine Sollbesetzung ermittelt. Unter Zugrundelegen der unterschiedlichen erwarteten Ausschöpfungen der einzelnen Stichproben und Erhebungsmethoden wird dann der Umfang der Erhebungsstichprobe errechnet.

2.3 Die Teilstichproben des IAB-Betriebspanels

Der Längsschnittcharakter des IAB-Betriebspanels spiegelt sich auch in der Stichprobe wider. Zum einen wird versucht, möglichst viele teilnehmende Betriebe über einen längeren Zeitraum hinweg zu befragen. Zum anderen muss die Dynamik von Betriebsschließungen und „neuen“ Betrieben auch in der Stichprobe des IAB-Betriebspanels abgebildet werden. Deshalb setzt sich die jährliche Bruttostrichprobe aus jeweils vier Teilstichproben zusammen:

1. Antwortbetriebe des Vorjahres („Wiederholer“),

⁸ Die Veränderung der Wirtschaftszweig-Systematik von WZ93 auf WZ2003 (Statistisches Bundesamt: 2003) schränkt die Vergleichbarkeit von Branchen nicht ein, da die Veränderungen unterhalb der im IAB-Betriebspanel genutzten Klassifikation stattgefunden haben.

⁹ Eine Ausnahme gilt hier für die Aufstockung für das Verarbeitende Gewerbe Ostdeutschlands. Auf Wunsch des IW Halle wurde hier nicht West-Berlin hinzugenommen, sondern die Betriebe aus Ost-Berlin nicht mehr einbezogen. Die Aufstockungsstichprobe für das Verarbeitende Gewerbe bezieht sich ab 2007 also auf alle ostdeutschen Bundesländer ohne Berlin.

2. erneut befragbare Ausfälle des Vorjahres („Nachbearbeitungsstichprobe“),
3. „neue“ Betriebsnummern,
4. Ersatzstichprobe.

Diese Teilstichproben sind notwendig, um Kontinuität und Wandel in der Betriebslandschaft abzubilden. Teilstichprobe 1, die Antworterbetriebe aus dem Vorjahr, bilden die Betriebe ab, die von einem auf das nächste Jahr zum Bestand der Betriebe gehören. Durch sie wird der Längsschnittcharakter des IAB-Betriebspanels bestimmt. Die erneut befragbaren Ausfälle des Vorjahres (Teilstichprobe 2) erhöhen die Fallzahlen im Querschnitt. Bei einer länger laufenden Betriebsbefragung wie dem IAB-Betriebspanel würde eine Konzentration nur auf die Betriebe, die von einem auf das andere Jahr zum Bestand gehören, zu Selektionseffekten führen. Über einen längeren Zeitraum hinweg bestehende Betriebe, unterscheiden sich in vielen betrieblichen Charakteristika von neu gegründeten. Um diese Dynamik ebenfalls abzubilden, wird die Stichprobe des IAB-Betriebspanels jedes Jahr um „neue“ Betriebsnummern (Teilstichprobe 3) ergänzt. Dabei handelt es sich um Betriebsnummern, die zum Stichtag mindestens einen sozialversicherungspflichtig Beschäftigten hatten, im Jahr zuvor jedoch nicht. Ein solcher Betrieb muss nicht zwangsläufig – wie vorher bei der Beschreibung der Vergabe von Betriebsnummern dargestellt – neu gegründet sein. Es kann sich auch um eine „ruhende“ Betriebsnummer handeln oder um einen Betrieb, der schon länger besteht, aber erst seit kurzem einen sozialversicherungspflichtig Beschäftigten hat. Diese Definition weicht von der der Neugründungen im Sinne von neu entstandenen Betrieben ab. Diese Betriebe sollten also nicht unkritisch als Neugründungen identifiziert werden. Hierzu sollten zusätzlich die Angaben aus dem Fragebogen herangezogen werden. Im Folgenden wird das Attribut „neu“ für diese Betriebsnummern immer in Anführungszeichen geschrieben, um falsche Schlussfolgerung zu vermeiden. Trotz dieser Schwierigkeiten, die sich aus der Systematik der Betriebsdatei ergeben, ist diese Vorgehensweise die einzige Möglichkeit, um das Gründungsgeschehen von Betrieben im Rahmen der Stichprobenziehung abzubilden.

Zusätzlich zu den eben beschriebenen Stichproben ist es unter Umständen notwendig, weitere Betriebe aus dem Bestand zum Stichtag neu in die Stichprobe aufzunehmen, um Ausfälle auszugleichen und die erforderliche Fallzahl in den einzelnen Bundesländern zu erreichen (Teilstichprobe 4).

Bis zur Welle des Jahres 2001 waren alle Betriebsnummern, die jemals in einer der Wellen in der Bruttostichprobe enthalten waren und nicht oder nicht mehr an der Befragung teilnehmen, für alle weiteren Wellen gesperrt. Das heißt, sie standen nicht mehr für die Stichprobenziehung zur Verfügung. Ab der Welle 2002 können Betriebsnummern, die bereits einmal zur Bruttostichprobe gehörten, jedoch zwischenzeitlich zu Ausfällen wurden, nach einer Karenzzeit

erneut gezogen werden. Grund für diesen Entschluss war, dass die Grundgesamtheit in manchen Branchen oder Bundesländern in den oberen Betriebsgrößenklassen fast ausgeschöpft war und somit Probleme bei der Besetzung von Zellen in der Schichtungsmatrix aufgetreten wären. Die Karenzzeit wird in Abhängigkeit vom Aufstockungsbedarf bestimmt. Für die Verarbeitung im Datensatz wird eine systemfreie Identifikationsnummer (idnum) vergeben. Aus organisatorischen Gründen erhalten die Betriebe, die erneut in die Bruttostichprobe gezogen wurden, eine neue Identifikationsnummer.¹⁰

¹⁰ Dabei kann nicht davon ausgegangen werden, dass ein Betrieb mit unterschiedlichen Identifikationsnummern und gleicher Betriebsnummer in den unterschiedlichen Wellen die gleiche wirtschaftliche Einheit widerspiegelt (vgl. Kap. 2.1 und Kap. 3).

3. Erhebungsdesign, Feldarbeit und Editing

Das Wichtigste in Kürze

Der Fragebogen setzt sich zusammen aus einem Block an Fragen, die in jedem Jahr identisch erhoben werden. Dazu kommen Themenbereiche, die in regelmäßigen Abständen wiederholt werden. Bei Längsschnittanalysen ist auf die genaue Formulierung der Fragen zu achten, da Abweichungen in den Frageformulierungen vorkommen. Zusätzlich werden jedes Jahr wechselnde aktuelle Schwerpunktthemen erfragt.

Die Erhebung erfolgt in der Regel in Form von persönlich-mündlichen Interviews, in einigen Bundesländern werden auch schriftlich-postalische Befragungen durchgeführt. Entsprechend sollte bei Analysen der betroffenen Bundesländer das Merkmal der Befragungsart berücksichtigt werden.

Die Feldphase findet jährlich im dritten Quartal statt. Die befragten Betriebe werden im Vorfeld durch Ankündigungsschreiben informiert. Vor Beginn des Interviews wird die richtige Betriebseinheit identifiziert. Dies wird in erster Linie durch einen Abgleich der Beschäftigtenzahlen des befragten Betriebs mit den Angaben aus der Vorjahresbefragung bzw. aus der BA-Statistik erreicht.

Insgesamt erzielt das IAB-Betriebspanel für Befragungen dieser Art gute Rücklaufquoten. Allerdings gibt es in Abhängigkeit der Stichprobenzugehörigkeit und Erhebungsmethode Unterschiede: Wiederholerbetriebe haben eine höhere Ausschöpfung als Erstbefragte und persönlich-mündliche befragte Betriebe haben eine höhere Rücklaufquote als schriftlich befragte Betriebe.

Um die Datenqualität zu verbessern, werden die Daten mit Hilfe eines aufwändigen Editings überprüft und Fehler korrigiert.

3.1 Fragebogenkonzeption und Pretest

Neben der Stichprobe muss auch der Fragebogen dem Längsschnittcharakter des IAB-Betriebspanels Rechnung tragen. Dabei geht es vor allem darum, bestimmte Informationen regelmäßig zu erheben, um Entwicklungen zu messen. Gleichzeitig müssen aktuelle, arbeitsmarktpolitisch relevante Fragestellungen in die Befragung einfließen.

Der Fragebogen des IAB-Betriebspanels enthält zahlreiche Fragen, die in jeder Welle gestellt werden, um Veränderungen über die Zeit konsistent abzubilden. Dieses Basisprogramm an Fragen ist in der Regel über die Jahre gleich. Einige Fragen mussten jedoch im Laufe der Zeit

geändert werden. Diese Veränderungen müssen bei Längsschnittanalysen immer berücksichtigt werden.¹¹

Bis zur Welle 2007 gehörten zu diesem Basisprogramm die Themenblöcke Beschäftigungsentwicklung, Geschäftspolitik und -entwicklung, Ausbildung, Personalstruktur und Personalbewegungen, Investitionen, Löhne und Gehälter sowie Tarifbindung. Darüber hinaus sind bestimmte Themenblöcke in größeren Abständen regelmäßig im Fragebogen, z.B. Themen wie Weiterbildung, Arbeitszeit, öffentliche Förderung und Innovationen. Zu beachten ist zudem, dass die „allgemeinen Angaben“ im Fragebogen nicht in allen Jahren für alle Betriebe erhoben werden, sondern meist nur für die erstbefragten Betriebe. Für wiederholt befragte Betriebe werden die entsprechenden Angaben aus früheren Wellen fortgeschrieben.

Im Jahr 2007 wurde die Fragebogenkonzeption etwas verändert. Das Basisprogramm wurde durch Basisindikatoren aus den bislang mehrjährigen Themenblöcken Weiterbildung, Innovation und Arbeitszeit ergänzt. Diese werden ab der Erhebungswelle 2008 jährlich abgefragt.¹² Für alle regelmäßig im IAB-Betriebspanel abgefragten Themenblöcke sind nun die Basisinformationen in jedem Jahr vorhanden, wodurch sich das Analysepotenzial erhöht, da inhaltliche Verknüpfungen auch im Querschnitt möglich sind. In zweijährigem Abstand werden diese Basisindikatoren durch zusätzliche Fragen ergänzt oder vertieft.¹³

Neben diesen festen Bestandteilen des Fragebogens werden jährlich wechselnde aktuelle Schwerpunktthemen erfragt. Dazu gehören beispielsweise Fragen zum Fachkräftebedarf, zur Beschäftigung von Älteren oder Beschäftigungs- und Standortsicherung.

Das Design des IAB-Betriebspanels wurde Anfang der 90er Jahre entwickelt und umfangreich getestet. Vor der Ausweitung der Befragung auf die ostdeutschen Länder wurde von TNS Infratest und dem Institut für sozioökonomische Strukturanalysen (SÖSTRA) im Auftrag des IAB Testerhebungen durchgeführt. Des Weiteren werden die für aktuelle Themenschwerpunkte neu entwickelten oder umformulierten Fragen im Pretest getestet. Bundesweit finden in ca. 100 Betrieben Interviews mit diesen neuen Fragen statt. Die speziell geschulten Interviewerinnen und Interviewer¹⁴ testen dabei nicht nur die reale Interviewsituation, sondern gehen insbesondere auf

¹¹ Eine Beschreibung für zeitkonsistente Anpassungen von Beispielvariablen findet sich im FDZ Datenreport 2/2006 (Alda et al. 2006). Darüber hinaus wird es mit PanelWhiz in Kürze ein Stata-Werkzeug geben, mit dessen Hilfe Paneldatensätze einfach und schnell erstellt werden können (weitere Informationen zu PanelWhiz: <http://www.panelwhiz.eu/>).

¹² Zur Umsetzung dieses Konzepts wurden einzelne Fragen aus dem Basisprogramm gestrichen. Dazu gehören unter anderem die Fragen nach den geplanten Investitionen bzw. die Fragen nach Investitionen in Informations- und Kommunikationstechnik.

¹³ Damit wurde ein gleichmäßiger Rhythmus für die mehrjährigen Module geschaffen. Zuvor wurden manche Themenblöcke nur alle drei Jahre erhoben.

¹⁴ Dabei handelt es sich um die Projektmitarbeiterinnen und -mitarbeiter des IAB in den Stützpunktagenturen (ProlAB).

Verständnis- und Interpretationsprobleme bei den Interviewten ein. Die so gewonnen Erkenntnisse fließen in die Entwicklung des Fragebogens ein.

3.2 Erhebungsmethode

Die Befragung erfolgt in der Regel in Form von persönlich-mündlichen Interviews, die von Mitarbeiterinnen und Mitarbeitern von TNS Infratest in den Betrieben durchgeführt werden. Entscheidend für den Erfolg der Befragung ist die Kontinuität bei den Interviewerinnen und Interviewern. Daher wird darauf geachtet, dass ein Betrieb möglichst immer von derselben Person befragt wird. Da im Fragebogen relativ viele Fragen zu Zahlenangaben enthalten sind, die die Auskunftspersonen nicht spontan beantworten können, kann der Fragebogen im Betrieb hinterlassen werden. Auf diese Weise können Angaben, die beim Interview nicht verfügbar sind, im Anschluss recherchiert und ergänzt werden. Diese Option wird vor allem von größeren Betrieben genutzt. Die Mehrheit der Betriebe wird allerdings vollständig persönlich-mündlich befragt (2006: 73%).

Seit 2000 werden teilweise schriftliche Befragungen vorgenommen. Dies wurde notwendig, weil in einigen Bundesländern Aufstockungsstichproben realisiert werden sollten, um Auswertungen auf Landesebene durchführen zu können. Die finanziellen Mittel für die große Anzahl von persönlichen Interviews reichten jedoch nicht aus, sodass in diesen Bundesländern ein Großteil der Stichprobe schriftlich-postalisch befragt wird. Betroffen davon waren Bremen von 2000 bis 2002 und Hamburg 2000 und 2001. Im Saarland werden seit 2001 und in Schleswig-Holstein seit 2002 die Betriebe aus den Aufstockungsstichproben nach wie vor schriftlich-postalisch befragt. Für die ersten postalischen Befragungen des Jahres 2000 wurde getestet, inwieweit sich durch den Methodenmix inhaltliche Verzerrungen ergaben. Dabei stellten sich teilweise signifikante Unterschiede im Antwortverhalten zwischen den persönlich-mündlich und den schriftlich-postalischen Interviews heraus. Bei den Analysen der betroffenen Bundesländer in den entsprechenden Jahren sollte daher ggf. das Merkmal der Befragungsart berücksichtigt werden.

3.3 Feldphase und Identifikation der richtigen Betriebseinheit

Da die Befragungen im dritten Quartal statt finden, beziehen sich die Fragen zu Bestandsgrößen in der Regel auf den Stichtag 30.6. bzw. zu Stromgrößen (wie z. B. Personaleinstellungen oder –abgänge) auf die erste Hälfte des Befragungsjahres. Damit sollen Erinnerungsfehler minimiert werden. Mitte Juni werden die Betriebe mit Ankündigungsschreiben¹⁵ auf die Befragung aufmerksam gemacht. Anschließend nehmen die Interviewerinnen und Interviewer Kontakt zu den Betrieben auf. Die schriftlich befragten Betriebe erhalten den Fragebogen mit den Anschreiben Ende Juni. Mitte Juli und Mitte August werden diese Betriebe anhand von Erinnerungsschreiben

¹⁵ Die Betriebe erhalten Anschreiben der Bundesagentur für Arbeit (BA) und der Bundesvereinigung der Deutschen Arbeitgeberverbände (BDA) mit der Bitte, sich an der Befragung zu beteiligen.

ggf. nochmals zur Teilnahme motiviert. Sowohl die persönlich-mündlichen als auch die schriftlich-postalischen Erhebungen sind in der Regel bis Mitte Oktober abgeschlossen. Parallel dazu werden die erfassten Daten bereits geprüft und Fehler bereinigt (siehe Kapitel 3.5).

Vor Beginn des Interviews müssen die Interviewerinnen und Interviewer sicherstellen, dass das Interview für die richtige Betriebseinheit durchgeführt wird. Für Längsschnittanalysen ist es von grundlegender Bedeutung, dass sich die Interviews immer auf dieselbe Betriebseinheit beziehen. Nur so kann sichergestellt werden, dass Veränderungen im Zeitverlauf (wie zum Beispiel die Beschäftigungsentwicklung) reale Veränderungen abbilden und nicht auf Abweichungen der befragten Betriebseinheit zurückzuführen sind. Die Identifikation der richtigen Betriebseinheit geschieht zu Beginn des Interviews anhand des Betriebsnamens, der Anzahl der sozialversicherungspflichtig Beschäftigten zum Stichtag (30.6. des Vorjahres) und – bei Abweichungen – anhand der Betriebsnummer.

Bei erstmals zu befragenden Betrieben wird der Abgleich der Beschäftigtenzahlen mit der Anzahl der sozialversicherungspflichtig Beschäftigten zum Stichtag (30.6. des Vorjahres) laut Betriebsdatei vollzogen. Ist es nicht möglich, das Interview mit genau der gezogenen Einheit durchzuführen, weil diese keine wirtschaftlich sinnvolle Einheit darstellt oder im Betrieb keine gesonderten Informationen speziell für diese Einheit vorliegen, kann für die Befragung nach festgelegten Regeln von der ursprünglich ausgewählten Einheit abgewichen werden. Die befragte Betriebseinheit muss aber einen Bezug zu der gezogenen Betriebsnummer haben. Dies kann sein, indem eine größere Betriebseinheit befragt wird, in der die Betriebsnummer enthalten ist, oder indem eine kleinere Betriebseinheit befragt wird, für die die Betriebsnummer in Teilen gilt. Diese Abweichungen werden im Adressprotokoll dokumentiert, so dass im Folgejahr dieselbe (abweichende) Einheit wieder befragt werden kann. Ist bei erstmals zu befragenden Betrieben kein Bezug zur gezogenen Betriebsnummer herzustellen, wird kein Interview durchgeführt.

Bei wiederholt befragten Betrieben wird die richtige Betriebseinheit anhand der Beschäftigtenzahlen aus dem Vorjahresfragebogen ermittelt. Treten Abweichungen auf, wird auch hier anhand der Betriebsnummer festgestellt, ob zumindest ein Teil des Betriebs bzw. eine übergeordnete Einheit die entsprechende Betriebsnummer trägt. Ist dies der Fall, wird die Befragung durchgeführt und das Verhältnis von befragter Einheit und Betriebsnummer auf dem Adressprotokoll dokumentiert. Diese Betriebe können dann für die Querschnittserhebung ausgewertet werden, für Panelanalysen stehen sie jedoch nicht mehr zur Verfügung.

Die schriftlich befragten Betriebe erhalten mit dem Fragebogen die entsprechenden Informationen zur Bestimmung der richtigen Befragungseinheit und werden gebeten, alle Angaben auf diese Einheit zu beziehen.

3.4 Ausfälle bei Interviews und Fragen

Die Rücklaufquote lag in den Erhebungen jeweils zwischen 63% und 73%.¹⁶ Die Schwankungen in den Rücklaufquoten sind in erster Linie auf die unterschiedlich großen Aufstockungsstichproben zurückzuführen. Da bei den erstmals befragten Betrieben die Rücklaufquoten erheblich niedriger liegen als bei den Wiederholerbetrieben¹⁷, sinkt die Gesamtrücklaufquote in den Jahren mit hohem Aufstockungsanteil deutlich. Neben Unterschieden zwischen erstbefragten und wiederholt befragten Betrieben unterscheidet sich die Ausschöpfung auch in Abhängigkeit von der Erhebungsmethode. Die Teilnahmebereitschaft liegt bei der persönlich-mündlichen Befragung deutlich höher als bei den schriftlich befragten Betrieben. So lag z. B. 2006 die Rücklaufquote bei schriftlich-postalischen Erstbefragten mit 12,9% deutlich niedriger als bei den persönlich-mündlich Erstbefragten mit 36,3%. Auch bei den wiederholt befragten Betrieben lag 2006 die Ausschöpfung der schriftlichen Befragungen mit 61,9% niedriger als die Ausschöpfung bei den Interviews vor Ort. Für die persönlich-mündlich befragten Wiederholerbetriebe liegen dagegen die Rücklaufquoten stabil zwischen 81% und 84%.

Zur Beurteilung der Erhebungsqualität müssen neben den kompletten Ausfällen bei der Befragung auch die fehlenden Angaben einzelner Fragen betrachtet werden. Fragen mit hohem Nonresponse sind entweder schwer verständlich, schwer beantwortbar oder ihre Beantwortung wird häufiger abgelehnt. Diese Ausprägungen werden als „keine Angabe“ erfasst und mit „-9“ kodiert.¹⁸ Über die Wellen des IAB-Betriebspanels hinweg sind die Ausfallquoten für die betrieblich sensiblen Merkmale wie Geschäftsvolumen, Lohn- und Gehaltssumme, Anteil der Vorleistungen und Fremdkosten am Umsatz und Summe der Investitionszuschüsse immer am höchsten, jedoch relativ stabil.

Bei der schriftlichen Befragung liegen die Anteile für „keine Angabe“ deutlich höher als bei den mündlichen Interviews.

Die geringeren Anteile von Item-Nonresponse und die höheren Ausschöpfungsquoten bei den persönlich-mündlichen Interviews unterstreichen die Datenqualität bei der angewandten Erhebungsmethode.

¹⁶ Bei dieser Berechnung sind nicht auswertbare Interviews bereits abgezogen.

¹⁷ Z.B. lag die Rücklaufquote der Ergänzungsstichprobe 2006 bei 30%.

¹⁸ Für wenige Fragen in den ersten Wellen gibt es zusätzlich die Ausprägung „weiß nicht“ (-8). Da im IAB-Betriebspanel fast ausschließlich Fakten, die dem Betrieb bekannt sind, abgefragt werden, ist eine Unterscheidung von „weiß nicht“ und „keine Angabe“ meist nicht sinnvoll. Bei Fragen, die aufgrund der Filterführung im Fragebogen bestimmten Betrieben nicht gestellt werden, finden sich im Datensatz Systemmissings.

3.5 Datenprüfung (Editing)

Parallel zur Feldphase findet eine umfassende Prüfung der Daten statt. Neben der Kontrolle der Interviewerinnen und Interviewer wird dabei vor allem die Datenqualität optimiert. Um dies zu erreichen, werden die Angaben zu einzelnen Fragen auf ihre Konsistenz und Plausibilität hin geprüft. Zusätzlich werden hierfür sowohl Fragen, die mit der zu Prüfenden in Zusammenhang stehen, als auch Fragen des Vorjahres, soweit sie verfügbar sind, herangezogen. Lassen sich die Fehler/Unplausibilitäten nicht anhand der Fragebogenangaben klären, wird der Betrieb telefonisch kontaktiert und das Problem gemeinsam mit der Auskunftsperson gelöst. Im Rahmen des Editings wird nochmals der Bezug zur richtigen Betriebseinheit kontrolliert. Stellt sich im Editing heraus, dass bei erstbefragten Betrieben kein Bezug zur Betriebsnummer hergestellt werden kann und die Betriebe somit fälschlicherweise befragt wurden, werden diese Interviews im Nachhinein ausgeschlossen.¹⁹ Das gleiche gilt für Interviews mit hohen Anteilen fehlender bzw. fehlerhafter Angaben und Fragebögen mit Verdacht auf Interviewerfälschung.

¹⁹ Bei Wiederholerbetrieben werden – wie oben beschrieben - Abweichungen zugelassen. Die Abweichungen müssen plausibel sein und dokumentiert werden. Die Interviews werden für den Querschnitt herangezogen, nicht jedoch für den Längsschnitt.

4. Auswertungen im Querschnitt

Das Wichtigste in Kürze

Auswertungen im Querschnitt ermöglichen inhaltliche Analysen für eine Welle oder Zeitreihenvergleiche über mehrere Wellen.

Um für die Grundgesamtheit repräsentative Aussagen machen zu können, muss, wegen des disproportionalen Stichprobenansatzes, für deskriptive Auswertungen im Querschnitt hochgerechnet werden. Durch diese Hochrechnung sind regionale, branchen- und größendifferenzierte Auswertungen möglich. Unterhalb der bei der Hochrechnung berücksichtigten Dimensionen der Schichtungsmatrix sind valide Aussagen nur eingeschränkt möglich.

Basis für die Auswertungen im Querschnitt sind die Querschnittsfälle der jeweiligen Welle. Das sind alle Betriebe mit mindestens einem sozialversicherungspflichtig Beschäftigten zum Stichtag (30.6. des Vorjahres). Diese Fälle können mit der Variablen `querxxxx` ausgewählt werden. `xxxx` steht für das jeweilige Erhebungsjahr (z.B. `quer2006` für das Erhebungsjahr 2006).

Der jeweilige Querschnittshochrechnungsfaktor ist mit `hrxxxxq` (z.B. `hr2006q`) bezeichnet.

Da das IAB-Betriebspanel auf einer Stichprobe basiert, müssen Fehlertoleranzen bei Auswertungen berücksichtigt werden.

Auch bei multivariaten Auswertungen ist der Disproportionalität der Stichprobe in geeigneter Weise Rechnung zu tragen.

Das Erhebungsdesign des IAB-Betriebspanels ermöglicht Auswertungen im Querschnitt und im Längsschnitt. Für beide Auswertungsmöglichkeiten ist zu beachten, dass die jeweils richtigen Fälle für die Auswertung ausgewählt werden und die Hochrechnung entsprechend angewendet wird. Die Auswertungen folgen jeweils einer unterschiedlichen Logik: Im Querschnitt werden die Betriebe betrachtet, die in dem relevanten Jahr zum Bestand gehörten. Im Längsschnitt hingegen können betriebsindividuelle Entwicklungen von Betrieben beobachtet werden, die über einen längeren Zeitraum hinweg bestanden haben. Zudem kann die Gründungs- und Schließungsdynamik²⁰ mit in die Analyse aufgenommen werden. Sowohl die Logik der Definition von Querschnitts- und Längsschnittsfällen also auch die jeweilige Hochrechnung unterscheiden sich voneinander. Da in den meisten Fällen entweder eine Querschnitts- oder eine Längsschnittsanalyse durchgeführt wird, werden im Folgenden die beiden Ebenen getrennt voneinander dargestellt.

4.1 Definition der Querschnittsfälle

Die Querschnittsfälle repräsentieren die Grundgesamtheit der jeweiligen Welle. Die Kriterien für einen Querschnittsfall sind, dass eine gültige Befragung vorliegt und der Betrieb zum Stichtag, 30.6. des Vorjahres, mindestens einen soziaversicherungspflichtig Beschäftigten hat. In den Datensätzen der einzelnen Wellen können die Querschnittsfälle jeweils über die String-Variable `querxxxx` (xxxx beschreibt die Jahreszahl der jeweiligen Welle, z.B. `quer2006` für die Welle 2006) identifiziert werden. Sie enthalten den Wert „Q“²¹. Diese Fälle bilden in jeder Welle die Basis für die Querschnittshochrechnung.

4.2 Querschnittshochrechnung

Die mehrfach disproportionale Struktur der Stichprobe wird mit Hilfe eines Gewichtungsverfahrens korrigiert und an die Struktur der Grundgesamtheit angepasst. Die Gewichtung erfolgt in Form einer Hochrechnung auf die Anzahl der Betriebe in der Grundgesamtheit. Dies gilt für Gesamt-, Ost- und Westdeutschland, aber auch für die einzelnen Bundesländer und die Betriebe des Verarbeitenden Gewerbes in Ostdeutschland.

Die Soll-Strukturen für die Hochrechnung des Querschnitts werden aus der Betriebsdatei der BA zum Stichtag (30.6. des Vorjahres) erstellt. Es handelt sich dabei um die Grundgesamtheit pro Bundesland über die Branchen- und Größenklasseneinteilung gemäß der Schichtungsmatrix für die Stichprobenziehung (ab 2003: 170-Zellen-Matrix). Da für eine valide Hochrechnung ausreichend Nettofälle pro Hochrechnungszelle notwendig sind, können nicht alle Bundesländer mit dieser 170-Zellen-Matrix hochgerechnet werden. In den Bundesländern mit einer geringeren Fallzahl erfolgt die Hochrechnung auf die jeweiligen Randverteilungen, also auf den Rand der Betriebsgrößenklassen und den Rand der Branchen.²²

In die Hochrechnung gehen alle Querschnittsfälle mit den erhobenen Ist-Daten der Beschäftigung, der Branchenzugehörigkeit und des Bundeslandes ein. Die Betriebe werden also nach ihren Angaben aus dem Fragebogen den einzelnen Zellen der Schichtungsmatrix zugeordnet. Daher kann es vorkommen, dass sich die Ziehungszelle und die Hochrechnungszelle eines Betriebes unterscheiden.²³

²⁰ Die Definition von „neuen“ Betrieben im IAB-Betriebspanel findet sich in Kap. 2.2. Zusätzlich können tatsächlich neu gegründete Betriebe über das bei der Befragung erhobene Gründungsjahr identifiziert werden.

²¹ In der Welle 1996 muss für die Definition zusätzlich noch der Buchstabe ‚L‘ mit herangezogen werden, der einmalige Zusatzfälle in West-Berlin beschreibt.

²² Das trifft zum Beispiel auf Bremen oder Schleswig-Holstein zu.

²³ Es kann vorkommen, dass einer Zelle im Verlauf der Erhebung mehr Betriebe zugeordnet werden als in der Grundgesamtheit in dieser Zelle vorhanden sind. Bis zum Jahr 1998 entstanden in solchen Fällen Hochrechnungsfaktoren kleiner eins. Ab 1999 liegt auch in solchen Fällen der kleinstmögliche Wert des Hochrechnungsfaktors bei eins.

Im ersten Schritt der Querschnittshochrechnung werden die „neuen“ Betriebsnummern der Ergänzungsstichprobe auf die entsprechende Sollvorgabe aus der Betriebsdatei hochgerechnet.

In einem zweiten Schritt werden alle Betriebe auf die Sollstruktur pro Bundesland laut BA-Betriebsdatei hochgerechnet. Dies erfolgt – wie bereits erwähnt – in den großen Bundesländern mit ausreichend Fallzahlen auf der Basis der Schichtungsmatrix, in den kleinen Bundesländern auf der Basis von Randverteilungen über Branche und Betriebsgröße. Die Hochrechnung der Betriebe auf die Sollstruktur laut BA kann problematisch sein, wenn in den einzelnen Zellen in der Stichprobe gar keine oder nur sehr wenig Fälle vorhanden sind. Um die Anzahl der Betriebe trotzdem richtig zu treffen, werden in diesem Fall Zellen zusammengefasst.

Für die Aufstockungsstichprobe im Verarbeitenden Gewerbe in Ostdeutschland wird eine gesonderte Hochrechnung durchgeführt. Hier erfolgt die Berechnung der Hochrechnungsfaktoren über 10 Betriebsgrößenklassen und 15 Branchen des Verarbeitenden Gewerbes ohne Differenzierung nach Bundesländern.

Abschließend wird noch einmal die Randverteilung der Betriebe über Branche und Betriebsgrößenklasse kontrolliert und angepasst.

In einem letzten Schritt erfolgt die Anpassung der Hochrechnungsfaktoren für die Anzahl der sozialversicherungspflichtig Beschäftigten zum Stichtag. Bei diesem Schritt wird darauf geachtet, dass trotz der Anpassung der Zahl der sozialversicherungspflichtig Beschäftigten die Anzahl der Betriebe über Betriebsgrößenklassen, Branchen und Bundesländern erhalten bleibt. Deshalb erfolgt die Anpassung der sozialversicherungspflichtig Beschäftigten innerhalb der Schichtungszellen. Die Gewichtungsschritte werden iterativ wiederholt.

Bei diesem Verfahren können unter Umständen sehr große Hochrechnungsfaktoren entstehen. Dies gilt vor allem für kleine Betriebe, bei denen die Zahlen in der Grundgesamtheit relativ hoch und die Auswahlsätze bei der Stichprobenziehung relativ klein sind. Um die damit verbundene Ausreißerproblematik zu begrenzen, wird die Höhe des Hochrechnungsfaktors kontrolliert und auf 3.000 begrenzt.

Im Ergebnis erhält jeder Betrieb einen individuellen Hochrechnungsfaktor (siehe unten: „Exkurs: Veränderung von Hochrechnungsfaktoren“). Im Anschluss an die Hochrechnung werden die gewichteten Daten in Kooperation mit dem IAB und anderen an der Auswertung des IAB-Betriebspanels beteiligten Instituten²⁴ intensiv geprüft und mit externen Quellen verglichen, um ein höchst mögliches Maß an Validität zu erreichen.

Exkurs: Veränderung von Hochrechnungsfaktoren

Auf Grund der Anlage der Hochrechnung ist es die Regel, dass ein Betrieb in zwei aufeinander folgenden Wellen unterschiedliche Hochrechnungsfaktoren zugewiesen bekommt. Dies gilt sowohl für die Querschnitts- als auch für die Längsschnitts- Hochrechnung. Dies geschieht unter anderem, weil:

- sich die Sollstruktur (die Zahl der Betriebe oder Beschäftigten in der Sollstruktur der jeweiligen Hochrechnungszelle) ändert,
- sich die Zahl der befragten Betriebe in der Hochrechnungszelle auf Grund von Aufstockungen oder Ausfällen verändert und/oder
- ein Betrieb zwischen zwei Wellen die Größenklasse, die Branche oder das Bundesland wechselt.

4.3 Hinweise zur Auswertungen im Querschnitt

In den Datensätzen ist der jeweilige Querschnittshochrechnungsfaktor mit hrxxxxq (z.B. hr2006q) gekennzeichnet. Auf Basis der geschilderten Hochrechnung können alle im Fragebogen erhobenen Variablen hochgerechnet werden. Es ist – je nach Betrachtungszeitraum – möglich, regionale (nach Bundesländern), branchenspezifische und größendifferenzierte Auswertungen²⁵ durchzuführen. Dabei ist allerdings zu beachten, dass unterhalb der bei der Hochrechnung berücksichtigten Schichtungsdimensionen (z.B. auf Kreisebene) keine validen Aussagen möglich sind. Des Weiteren ist zu berücksichtigen, dass für Hamburg auf Grund zu geringer Fallzahlen bundeslandsspezifische Auswertungen nicht möglich sind. Bei stark differenzierten deskriptiven Analysen ist außerdem die Fallzahl zu beachten. Es wird empfohlen, dass jede Analysegruppe ungewichtet mit mindestens 20 Betrieben besetzt ist.

²⁴ Neben dem IAB wird das IAB-Betriebspanel im Auftrag der an der Erhebung beteiligten Bundesländer und Institutionen von weiteren Forschungseinrichtungen ausgewertet. Im Jahr 2006 waren dies das BAW Institut für regionale Wirtschaftsforschung GmbH, das Institut für Angewandte Wirtschaftsforschung (IAW) Tübingen, das Institut für Sozialökonomische Strukturanalysen GmbH (SÖSTRA), das Institut für Wirtschaft, Arbeit und Kultur Frankfurt (IWAK), das Institut für Wirtschaftsforschung Halle (IWH) und das Internationale Institut für empirische Sozialökonomie (INIFES).

²⁵ Eine – in der Regel nicht überprüfbare, aber bei einem solchen Studiendesign gemeinhin als gegeben angesehen – Voraussetzung dafür ist, dass für das betrachtete Merkmal eine Selektion hinsichtlich der Teilnahmebereitschaft und des Antwortverhaltens nur entlang der Hochrechnungsdimensionen vorliegt.

Wie bei allen Stichprobenerhebungen ist auch bei den Ergebnissen des IAB-Betriebspanels eine statistische Fehlertoleranz zu berücksichtigen. Die im Rahmen der Stichprobenerhebung gewonnenen Daten sind Schätzwerte für den Anteil bzw. die absolute Größe der entsprechenden Merkmale in der Grundgesamtheit und mit einer gewissen Unschärfe behaftet, die sich berechnen lässt. Generell gilt: Je größer die Streuung des betreffenden Merkmals in der Grundgesamtheit und je kleiner die der Schätzung zu Grunde liegende (ungewichtete) Fallzahl, desto größer ist der statistische Unschärfebereich. Für die Interpretation der Daten ist es wichtig, die Größe des Unschärfebereichs zu kennen.

Aus der Fehlertoleranztafel (siehe Tabelle H im Anhang) kann der Unschärfebereich für verschiedene Fallzahlen und Anteilswerte näherungsweise abgeschätzt werden. Bei Anteilswerten und einer Auswertung mit 15.000 Fällen, können als Faustregel Abweichungen von 2 Prozentpunkten und mehr als Unterschied interpretiert werden. Einigen Fragen wurden im Fragebogen Filter vorgeschaltet, wodurch sich die zugrunde liegende Fallzahl verringert. Hier sind höhere Toleranzgrenzen anzusetzen. Auch bei Absolutzahlen ist die oben genannte Faustregel nicht anwendbar. Eventuell bietet Ihre Statistiksoftware eine einfache Möglichkeit zur Ausgabe von Konfidenzbereichen.²⁶ Ist die Verteilung eines Schätzers für eine Kennzahl unbekannt oder die Berechnung von Konfidenzintervallen für diese Kennzahl nicht in Ihrer Software implementiert, lassen sich diese auch anhand des Bootstrap-Verfahrens bestimmen (Green 2003, S. 924f).

Für deskriptive Auswertungen der Querschnittsdaten wird empfohlen, Beschäftigtenzahlen auf ganze Tausender zu runden und Prozentangaben ohne Nachkommastellen auszuweisen.

Aufgrund der disproportionalen Stichprobe, die durch die Hochrechnung korrigiert wird, können sich die Ergebnisse gewichteter und ungewichteter multivariater Schätzungen unterscheiden. Im Querschnitt können – nach Ermessen des Forschers/der Forscherin – sowohl gewichtete als auch ungewichtete multivariate Verfahren angewendet werden. Um die Effekte der disproportional geschichteten Stichprobe zu kontrollieren, wird empfohlen die jeweiligen Schichtungsvariablen (Branche, Betriebsgröße und Bundesland zum Stichtag des Vorjahres) als unabhängige Kontrollvariablen in die Modelle aufzunehmen.

²⁶ So erlaubt Stata die Schätzung hochgerechneter Fallzahlen einschließlich Konfidenzintervalle mit dem Befehl `-svy: total-`, Anteile werden mit `-svy: proportion-` berechnet. Neben dem Hochrechnungsfaktor ist dabei auch ein Identifikator der Zellen der Schichtungsmatrix zu übergeben.

5. Auswertungen im Längsschnitt

Das Wichtigste in Kürze

Auswertungen im Längsschnitt ermöglichen das Betrachten betriebsindividueller Entwicklungen und Zusammenhänge. Außerdem kann die Gründungs- und Schließungsdynamik mit in die Analysen einbezogen werden.

Im IAB-Betriebspanel sind fünf Längsschnitte definiert und zwar mit den Startjahren 1993, 1996, 2000, 2003 und 2007.

Panelfälle sind Betriebe aus dem Bestand des Startjahres, „neue“ Betriebe in den Folgewellen und im Panelzeitraum erloschene Betriebe. Als wesentliche Voraussetzung für die Definition der Panelfälle muss dieselbe Betriebseinheit wie im Vorjahr befragt worden sein. Diese Fälle sind im Datensatz mit den Variablen panyy_zz identifizierbar, wobei yy das Startjahr und zz den aktuellen Rand des Längsschnitts beschreibt (z.B. pan93_06 definiert die Panelfälle aus der ersten Welle 1993 bis zum aktuellen Rand 2006).

Diese Definition impliziert, dass nicht alle Betriebe eines Querschnitts in einem das jeweilige Querschnittsjahr einschließenden Längsschnitt enthalten sind. Das gilt für die Betriebe der Ersatzstichprobe, die aus dem Bestand gezogen worden sind und für Betriebe, bei denen eine abweichende Einheit befragt wurde. Nur in der Startwelle des jeweiligen Längsschnitts sind alle Betriebe auch Querschnittsfälle.

Der disproportionale Stichprobenansatz aus dem Querschnitt setzt sich im Längsschnitt fort. Aus diesem Grund müssen auch deskriptive Längsschnittanalysen gewichtet vorgenommen werden, um repräsentative Aussagen über die Grundgesamtheit machen zu können. Der jeweilige Hochrechnungsfaktor ist der Variable hryy_zzp (z.B. hr93_06p) zu entnehmen.

Auch bei multivariaten oder panelökonometrischen Auswertungen ist der Disproportionalität der Stichprobe in geeigneter Weise Rechnung zu tragen und Fehlertoleranzen sind zu berücksichtigen.

An Hand des Wellenkennzeichens können darüber hinaus eigene Längsschnitte definiert werden. Für diese liegen allerdings keine Hochrechnungsfaktoren vor.

Anders als bei Analysen im Querschnitt bieten Längsschnittauswertungen die Möglichkeit, auf betriebsindividueller Ebene Entwicklungen und Zusammenhänge zwischen betrieblichen Merkmalen im Zeitverlauf zu analysieren. Sowohl die Definition von Panelfällen als auch die Längsschnitthochrechnung unterscheiden sich in der Konzeption deutlich von den entsprechenden Vorgängen im Querschnitt. Beim Längsschnitt muss die Dynamik von Gründung und

Schließung, sowie Veränderungen der einzelnen Betriebe im Zeitverlauf wie zum Beispiel das Wachsen und Schrumpfen von Betrieben mit berücksichtigt werden. Zur Veranschaulichung einer hochgerechneten Längsschnittsanalysen empfehlen wir das Kapitel 6.2 des IAB-Forschungsberichts 5/2007 (Fischer et al. 2007).

5.1 Definition von Längsschnittsfällen

Panelfälle sind Betriebe aus der Querschnittsstichprobe des jeweiligen Basisjahres sowie alle „neuen“ Betriebe aus den Ergänzungs- und Aufstockungsstichproben der Folgejahre. Für die Panelfälle muss seit der jeweiligen Erstbefragung bis zum aktuellen Rand entweder ein auswertbares Interview für immer dieselbe Betriebseinheit vorliegen oder die Information, dass die Betriebsnummer erloschen ist. Als Panelfälle werden also nicht nur die Betriebe definiert, die zu beiden Eckpunkten des Längsschnitts mindestens einen sozialversicherungspflichtig Beschäftigten haben. Würde man die Panelfälle auf diese Definition beschränken, wäre eine Analyse der Betriebsdynamiken im Zeitverlauf nur als Ausschnitt möglich. Zur betrieblichen Dynamik gehört das Erlöschen von Betrieben ebenso wie das Hinzukommen „neuer“ Betriebe. Um diese Dynamiken abbilden zu können, beinhaltet die Panelfalldefinition im Zeitraum erloschene Betriebe, „neue“ Betriebe und Betriebe ohne sozialversicherungspflichtig Beschäftigte. Betriebe, die ihren Standort von West- nach Ostdeutschland bzw. umgekehrt verlegen, werden als Panelfälle ausgeschlossen.

Ausgangspunkt für die Bildung eines Längsschnittes sind alle in diesem Startjahr existierenden Querschnittsfälle. Sie repräsentieren den Bestand aller Betriebe im Startjahr und für sie soll die Entwicklung in den nächsten Jahren beobachtet werden. In den Folgewellen werden Panelfälle dann folgendermaßen definiert:

1. Panelfälle aus der Vorwelle, die noch bestehen:

Diese Gruppe repräsentiert die sogenannten überlebenden Betriebe. Um Verzerrungen bei den Analysen zu vermeiden, werden aber nur die überlebenden Betriebe als Panelfälle weiter geführt, für die das Interview für dieselbe Betriebseinheit wie in der vorherigen Welle durchgeführt werden konnte. Ist es nicht möglich, für dieselbe Einheit wie im Vorjahr ein Interview zu machen, steht dieser Betrieb zwar für den Querschnitt, nicht aber für den Längsschnitt zur Verfügung.

2. Panelfälle aus der Vorwelle, die in der Vorwelle oder in früheren Wellen erloschen sind:

Diese Betriebe bilden das Erlöschen von Betrieben ab. Für diese Betriebe kann kein Interview durchgeführt werden. Das bedeutet, dass sie von dem Jahr an, in dem sie erloschen sind, nur noch mit der Information weiter geführt werden, dass diese Betriebe im für die Panelfalldefinition gültigen Zeitraum erloschen sind.

3. „Neue“ Betriebsnummern aus den Ergänzungsstichproben der Folgejahre:

Neben bestehenden und erloschenen Betrieben bilden „neue“ Betriebe den dritten Teil der Betriebsdynamik. Die „neuen“ Betriebsnummern, die in jeder Welle in die Stichprobe aufgenommen werden (siehe Kapitel 2), sind also ebenfalls Bestandteil der Panelfalldefinition, wenn ein Interview mit der von der BA gezogenen Betriebseinheit durchgeführt werden konnte.²⁷

Die Betriebe, die in jeder Welle als Ersatz für Ausfälle aus dem Bestand gezogen werden, werden nicht in die Panelfalldefinition aufgenommen, da die Entwicklung des Bestands an Betrieben nur vom jeweiligen Startjahr aus beobachtet wird.

Im IAB-Betriebspanel sind unterschiedliche Längsschnitte definiert, die mit fünf verschiedenen Startjahren beginnen:

Längsschnitt 1 (nur für West-Deutschland) im Zeitraum 1993 bis 2006

Längsschnitt 2 im Zeitraum 1996 bis 2006,

Längsschnitt 3 mit dem Startjahr 2000 bis zum aktuellen Rand,

Längsschnitt 4 mit dem Startjahr 2003 bis zum aktuellen Rand.

Längsschnitt 5 mit dem Startjahr 2007 bis zum aktuellen Rand.

Die Panelfälle sind in den jeweiligen Datensätzen mit den Variablen `panyy_zz` identifizierbar, wobei `yy` das Startjahr und `zz` jedes beliebige Folgejahr beschreibt (z.B. `pan93_06` definiert die Panelfälle aus der ersten Welle 1993 bis 2006). Die Variablen sind mit „P“ belegt, wenn der Fall zum entsprechenden Längsschnitt gehört. Für diese definierten Panelfälle sind auch Hochrechnungsfaktoren vorhanden. Für die Analyse müssen die jeweils zu untersuchenden Zeiträume aber nicht unbedingt identisch sein mit den Zeiträumen der Längsschnitte. Der Analysezeitraum muss aber vollständig in dem jeweiligen Längsschnittszeitraum enthalten sein. Beispielsweise ist mit dem Längsschnitt von 1996 bis 2006 auch der Zeitraum 1998 bis 2001 auswertbar.

5.2 Längsschnittshochrechnung

Ähnlich wie bei deskriptiven Auswertungen der Querschnittsdaten gilt bei deskriptiven Auswertungen von Längsschnittsdaten, dass sie ohne Hochrechnung wegen der disproportionalen Stichprobenanlage zu verzerrten Ergebnissen führen. Bei einer Längsschnittshochrechnung für

²⁷ Erstbefragte Betriebe aus der Teilstichprobe „neue“ Betriebsnummern, bei denen eine abweichende Einheit befragt wurde (siehe Kap. 3.3), werden nicht als „neue“ Betriebe definiert und daher auch nicht bei der Konstruktion des Längsschnitts berücksichtigt.

eine Betriebsbefragung mit disproportionaler Stichprobe sind verschiedene Anforderungen zu beachten:

Die Betriebe, die in den jeweiligen Wellen zum Betriebsbestand gehören, müssen im Rahmen der Längsschnittshochrechnung die Strukturen über Branche und Betriebsgröße zum Stichtag der jeweiligen Welle reproduzieren. Das bedeutet, dass diese Betriebe jeweils für jede einzelne zum Längsschnitt gehörende Welle auf die Sollmatrix der Betriebsdatei hochgerechnet werden müssen.²⁸ Da viele Betriebe in mehreren Wellen im Bestand sind, bedeutet das für die Hochrechnung, dass diese Betriebe in verschiedenen Wellen auf unterschiedliche Grundgesamtheiten hochgerechnet werden müssen, aber für den gesamten Längsschnitt nur ein Faktor pro Betrieb möglich ist.

Im Zeitverlauf wachsende bzw. schrumpfende Betriebe stellen eine besondere Herausforderung für die Längsschnittshochrechnung dar, sofern die Veränderungen zu einem Wechsel der für die Hochrechnung relevanten Größenklasse führen, da die durchschnittlichen Hochrechnungsfaktoren wegen der disproportionalen Stichprobenanlage in Abhängigkeit von der Betriebsgröße stark variieren. Bei Neudefinition eines Längsschnitts dient der Querschnittsfaktor der Basiswelle als Eingangsfaktor für die Längsschnittshochrechnung, bei bereits bestehenden Längsschnitten der Längsschnittfaktor der Vorwelle. Wechselt ein Betrieb zwischen zwei Wellen beispielsweise von einer Größenklasse in die nächst höhere, bekommt er – aufgrund der Tatsache, dass der Faktor der Vorwelle als Eingangsfaktor dient – zunächst einen zu hohen Eingangsfaktor, so dass dieser Betrieb ein zu hohes Gewicht in der Hochrechnung erhalten würde. Analoges gilt für den umgekehrten Fall: Schrumpft ein Betrieb von einer größeren Größenklasse in die nächst kleinere Klasse der Schichtungsmatrix, erhält dieser Betrieb zunächst einen zu kleinen Faktor, da er im Vorjahr aufgrund des Ausgleichs der disproportionalen Stichprobe als Großbetrieb einen kleineren Faktor bekommen hat. Diese Effekte werden kontrolliert, sodass Verzerrungen bei der Auswertung der jeweiligen Welle minimiert werden.

Die Längsschnittshochrechnung ist so angelegt, dass sich bei einer Auswertung der Panelfälle mit dem Längsschnittshochrechnungsfaktor – für vergleichbar definierte Teilgruppen – für alle Fragen aus den bisherigen Erhebungen in etwa dieselben Verteilungen ergeben wie bei den entsprechenden Querschnittauswertungen der betreffenden Welle.

Selbst bei einer so großen Stichprobe, wie sie im IAB-Betriebspanel zur Verfügung steht, sind bei der Längsschnittshochrechnung theoretisch erheblich mehr Parameter zu kontrollieren, als praktisch möglich und sinnvoll ist. Die Längsschnittshochrechnung muss sich daher auf die wesentlichen Parameter konzentrieren.

²⁸ Aufgrund der geringeren Fallzahlen im Längsschnitt wird die Branchengliederung bei der Längsschnittsgewichtung auf sieben Branchen zusammengefasst.

Die Längsschnittshochrechnung erfolgt in acht aufeinander aufbauenden Schritten, die im Anhang detailliert erläutert werden. Mit Ausnahme des Längsschnittes ab 1993, in dem nur Westbetriebe enthalten sind, wird in allen Schritten grundsätzlich eine Differenzierung nach Ost- und Westdeutschland vorgenommen.²⁹

In den einzelnen Schritten zur Berechnung der Längsschnitthochrechnungsfaktoren werden folgende Aspekte berücksichtigt:

- die Struktur der Betriebe nach Branche, Größenklasse und Bundesland am aktuellen Rand,
- die Anzahl der bestehenden, erloschenen und neu gegründeten Betriebe in jedem Jahr des den Längsschnitt umfassenden Zeitraums,
- das Überleben und Erlöschen der Betriebe aus dem Bestand des Startjahres sowie der im Verlauf des Längsschnitts hinzu gekommenen „neuen“ Betriebsnummern,
- das Wachsen und Schrumpfen der Betriebe aus dem Bestand des Startjahres sowie der im Verlauf des Längsschnitts hinzu gekommenen „neuen“ Betriebsnummern,
- disproportionale Ausfälle; Basis hierfür sind die Fragebogenangaben aus dem Vorjahr und
- die Branchen- und Größenstruktur der Grundgesamtheit in jeweils allen den Längsschnitt umfassenden Wellen.

5.3 Hinweise zur Auswertung im Längsschnitt

Der jeweilige Hochrechnungsfaktor ist der Variable `hryy_zzp` (z.B. `hr93_06p`) zu entnehmen. Auf Basis der geschilderten Hochrechnung können alle im Fragebogen erhobenen Variablen hochgerechnet werden. Es ist – je nach Betrachtungszeitraum – möglich, regionale (nach Bundesländern), branchenspezifische und größendifferenzierte Auswertungen durchzuführen. Bei den Analysen im Längsschnitt sind ebenso die Fehlertoleranzen und Fallzahlen bei differenzierten deskriptiven Analysen wie im Querschnitt (siehe Kap. 4.3) zu berücksichtigen. Im Anhang sind in der Tabelle F alle Panelfälle abgebildet und in Tabelle G die Panelfälle mit gültigem Interview in jedem Jahr.

²⁹ Wie im Kapitel 2.2 ausgeführt, gilt von 2007 an eine neue Definition von West- und Ostdeutschland. Demnach gehören zu Westdeutschland nur die westdeutschen Bundesländer ohne Berlin. Ostdeutschland umfasst alle ostdeutschen Bundesländer einschließlich Berlin.

Bei deskriptiven Längsschnittauswertungen ist zudem Folgendes zu beachten: Die Längsschnittshochrechnung ist mit größeren Unschärfen versehen als die Querschnittshochrechnung. Dies gilt vor allem für hochgerechnete Beschäftigtenzahlen. Es wird empfohlen, bei deskriptiven Längsschnittanalysen die Absolutzahlen sehr grob zu runden bzw. Verteilungsmaße anzuwenden. Hierbei ist zu beachten, dass die ermittelten Prozentwerte ebenfalls erhebliche Unschärfen aufweisen können.

Aufgrund der disproportionalen Stichprobe, die durch die Hochrechnung korrigiert wird, können sich die Ergebnisse gewichteter und ungewichteter multivariater Schätzungen unterscheiden. Im Längsschnitt können – nach Ermessen des Forschers/der Forscherin – sowohl gewichtete als auch ungewichtete multivariate Verfahren angewendet werden. Die meisten Nutzerinnen und Nutzer verwenden ungewichtete Daten. Um die Effekte der disproportional geschichteten Stichprobe zu kontrollieren, wird empfohlen die jeweiligen Schichtungsvariablen (Branche, Betriebsgröße und Bundesland zum Stichtag des Vorjahres) als unabhängige Kontrollvariablen in die Modelle aufzunehmen.

5.4 Das Wellenkennzeichen: Definition eigener Längsschnitte

Grundsätzlich besteht auch die Möglichkeit, eigene Panelfalldefinitionen für andere Zeiträume zu erstellen. Dies kann mit Hilfe des Wellenkennzeichens (wellxxxx; xxxx steht für das Erhebungsjahr) erfolgen. In dieser Variable werden die Antworten auf vier für die Identifikation von Längsschnittsfällen zentrale Fragen abgebildet:

1. Hat der Betrieb sozialversicherungspflichtige Beschäftigte am 30.6. des Vorjahres?
2. Liegt ein aktuelles, auswertbares Interview vor?
3. Hat der Betrieb bereits im Vorjahr an der Befragung teilgenommen?
4. Wurde dieselbe Betriebseinheit wie im Vorjahr befragt?

Jeder Betrieb der Bruttostichprobe wird für die jeweilige Welle eindeutig gekennzeichnet. Insgesamt umfasst diese String-Variable 11 Werte (siehe Tabelle 1).³⁰

³⁰ Bei der Vergabe des Wellenkennzeichens gelten folgende Regeln: Betriebe, die erstmalig in der jeweiligen Welle befragt werden und von denen ein auswertbares Interview vorliegt, erhalten den Code A. In der ersten Welle 1993 betrifft dies alle Fälle aus der Basisstichprobe. Ab der zweiten Welle gilt der Code A für alle Betriebe aus den Ergänzungs- und Aufstockungsstichproben. Ab der Welle 2002 ist es möglich, dass Betriebe erneut als Erstbefragte gelten und den Code A zugewiesen bekommen, wenn sie nach einem Ausfall erneut in die Ergänzungs- bzw. Aufstockungsstichprobe gezogen werden. Sie erhalten dann aber eine neue Identifikationsnummer. Die Vergabe des Codes A trifft sowohl auf die Fälle aus der Basisstichprobe West 1993 und Ost 1996, als auch auf alle „neuen“ Betriebsnummern aus den Ergänzungsstichproben in den jeweiligen Wellen und den Fällen aus den Aufstockungsstichproben zu. In der ersten Welle sind nur die Codes A und H erlaubt. Ab der zweiten Welle sind die Codes B, C, D, H und Y zulässig, wenn im Vorjahr der Code A bis G enthalten ist. Die Codes E und G erhalten nur Fälle, für die im Vorjahr der Code H vergeben wurde. Bei Betrieben, die in zwei aufeinanderfolgenden Wellen kein Interview gegeben haben, wird kein weiterer Interviewversuch angestrebt.

Tabelle 1: Wellenkennzeichen

		Kennbuchstaben	
		Mit	ohne
		sozialversicherungspfl. Beschäftigten am 30.6. des Vorjahres	
1.	Fälle mit Interview in lfd. Welle		
1.1.	Erstmals befragte Fälle	A	nicht zulässig
1.2.	wiederholt befragte Fälle		
1.2.1.	mit Interview im Vorjahr:		
	gleiche Einheit wie im Vorjahr	B	C
	andere Einheit wie im Vorjahr	D	nicht zulässig
1.2.2.	ohne Interview im Vor-Vorjahr		
	selbe Einheit wie im Vor-Vorjahr	E	nicht zulässig
	andere Einheit wie im Vor-Vorjahr	G	nicht zulässig
2.	Fälle ohne Interview in lfd. Welle		
2.1.	Ausfall, in Zukunft wieder befragbar		H
2.2.	Fälle aus Aufstockungen, die nicht mehr weiter befragt werden		W
2.3.	Ausfall, nicht mehr befragbar		X
2.4.	Betrieb erloschen:		
2.4.1.	in der lfd. Welle		Y
2.4.2.	bereits früher		Z

Möchte man einen eigenen Längsschnitt ³¹ nach der in Kapitel 5.1 vorgestellten Logik definieren, sind die Querschnittsfälle des Startjahres die Basis. Diese können über die Variable `querxxxx` ausgewählt werden. In den Folgejahren werden ausschließlich die Betriebe, für die das Wellenkennzeichen mit den Buchstaben B oder C (Wiederholerbetrieb, gleiche Einheit wie im Vorjahr wurde befragt), Y (erloschener Betrieb aus der laufenden Welle) oder Z (erloschener Betrieb aus der Vorwelle) als Panelfälle definiert. Hinzu kommen noch die „neuen“ Betriebs-

Diese Betriebe erhalten im ersten Ausfalljahr den Code H und im darauffolgenden Jahr den Code X. Definitiv erloschen Betriebe werden mit dem Kennbuchstaben Y gekennzeichnet, wenn sie in der laufenden Welle erloschen sind. In den darauffolgenden Wellen erhalten diese Fälle den Code Z. Betriebe, die mit den Codes W, X, Y und Z versehen sind, werden nicht weiter befragt und erhalten die gleiche Kennung in den darauffolgenden Welle. Nur beim Code Y wird der Kennbuchstabe Z in der Folgewelle fortgeschrieben.

³¹ Mit Hilfe der Wellenkennzeichen bietet sich auch eine alternative Möglichkeit zur Identifikation der Querschnittsfälle. Sie umfassen alle Betriebe mit den Kennbuchstaben A, B, D, E und G der jeweiligen Welle.

nummern, die sich aus der Variable neuxxxxr³² (xxxx steht für das Erhebungsjahr) identifizieren lassen. Bei der Definition von eigenen Längsschnitten ist darauf zu achten, dass stets dieselbe betriebliche Einheit befragt wurde. Beispiel: Ein Längsschnitt der sich auf die Wellen 1 bis 3 bezieht, sollte keine Fälle enthalten, bei der in der Welle 2 oder 3 das Wellenkennzeichen die Werte D oder G (andere Einheit befragt) annimmt. Es sei nochmals darauf hingewiesen, dass für selbst definierte Längsschnitte keine Hochrechnungsfaktoren zur Verfügung stehen. Deskriptive Auswertungen, die repräsentativ für die Grundgesamtheit sind, können somit nicht durchgeführt werden.

³² Diese Variable kann bei Bedarf dem Datensatz zugespielt werden. „Neue“ Betriebe sind mit „N“ gekennzeichnet, allerdings nur wenn keine abweichende Einheit befragt wurde (siehe Kap. 5.1). In den einzelnen Wellen befindet sich zusätzlich die Variable neuxxxx, die jedoch nur eine neue Betriebsnummer anzeigt. In den Wellen 1996 und 1997 sind Betriebe enthalten, die nur einmalig befragt wurden und somit nicht im Längsschnitt berücksichtigt werden können. Diese Information findet sich in der Variable zusatz (Ausprägungen W4 und W5), die ebenfalls zugespielt werden kann.

6. Datenzugang

Da den befragten Betrieben zugesichert wurde, dass ihre Daten nur in anonymisierter Form publiziert und nicht an Dritte weitergegeben werden, ist der Datenzugang für externe Wissenschaftlerinnen und Wissenschaftler nur über das Forschungsdatenzentrum (FDZ) der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB) möglich. Die Forschenden haben die Wahl zwischen einem Gastaufenthalt am FDZ oder der kontrollierten Datenfernverarbeitung.

Zum Beantragen der Datenfernverarbeitung muss für ein Projekt lediglich ein kurzes Anfrageformular ausgefüllt und an das FDZ gefaxt werden. Das FDZ entscheidet über die Genehmigung des Datenzugangs. In der Regel wird diese Genehmigung innerhalb weniger Tage per E-Mail versandt. Anschließend erstellen die Forschenden auf der Basis von Testdaten Auswertungsprogramme in SPSS, Stata oder SAS und schicken diese per E-Mail an das FDZ. Das FDZ führt die Auswertungen mit den Originaldaten durch und übermittelt die Ergebnisse nach einer Datenschutzprüfung zurück.

Um zum Gastaufenthalt an das FDZ zu kommen, ist im Vorfeld ein ausführlicher Antrag für das Forschungsprojekt zu stellen. Dieser wird vom FDZ und dem Bundesministerium für Arbeit und Soziales (BMAS) beurteilt. Nach der Genehmigung durch das BMAS wird ein Vertrag geschlossen. Anschließend wird ein Termin mit dem FDZ vereinbart. Im FDZ arbeiten die Gastwissenschaftler/innen an speziell eingerichteten Arbeitsplätzen direkt mit den Daten des IAB-Betriebspanels. Die Analyseergebnisse werden einer Datenschutzprüfung unterzogen, um sicherzustellen, dass nur absolut anonymisierte Ergebnisse das Haus verlassen.

Denjenigen, die bereits zum Gastaufenthalt am FDZ waren, steht es zudem offen, anschließend noch weitere Analysen für das Projekt über die Datenfernverarbeitung durchzuführen. Ein weiterer Antrag ist hierfür nicht nötig.

Über das FDZ kann auf alle Befragungswellen des IAB-Betriebspanels zugegriffen werden. Dabei wird für jede Welle ein eigener Datensatz mit allen Angaben aus dem Fragebogen, den Hochrechnungsfaktoren, Wellenkennzeichen und anderen technischen Variablen bereitgestellt. Zusätzlich zu diesen Datensätzen gibt es für jedes Jahr (ab 1998) einen sogenannten Organisationsfile. Hier werden die Identifikationsnummern aller bisher befragten Betriebe mit einigen zusätzlichen Informationen aus der BA-Statistik geführt. Nutzerinnen und Nutzer können einige dieser Merkmale zusätzlich erhalten, wenn sie den Datenbedarf ausführlich begründen. Das FDZ stellt folgende Merkmale aus dem Organisationsfile auf Antrag bereit: Kreiskennziffer, Gemeindegröße, Wirtschaftszweigklassifikation WZ1993 (für die Jahre 2000-2003) und Wirt-

schaftszweigklassifikation WZ2003 (ab 2004). Diese Merkmale sind in den Testdaten bereits enthalten.³³

Zur Vorbereitung der Analysen mit dem IAB-Betriebspanel bietet das FDZ eine Vielzahl von Arbeitshilfen online an, z.B. Datensatzbeschreibungen, Testdaten und Fragebögen. Ebenso werden auf der FDZ-Homepage die Datenzugangswege erläutert und alle Antragsformulare sind abrufbar. (<http://fdz.iab.de>)

Weitere Informationen zur aktuellen Befragungswelle, inhaltliche Ergebnisse und zahlreiche Literaturhinweise finden sich auf der Homepage des IAB-Betriebspanels (<http://betriebspanel.iab.de>).

³³ Zusätzlich können die Nutzerinnen und Nutzer die Variablen `zusatz` und `neuxxxxr` aus der Organisationsfile ohne Begründung erhalten. Diese Merkmale sind jedoch nicht in den Testdaten enthalten.

Literatur

- Alda, H./Bender, S./Gartner, H. (2005): The Linked Employer-Employee Dataset created from the IAB Establishment Panel and the Process-Produced Data of the IAB (LIAB). In: Schmollers Jahrbuch 125, 327-336.
- Alda, H./Dundler, A./Müller, D./Spengler, A. (2006): Aufbereitung eines Paneldatensatzes aus den Querschnittsdaten des IAB-Betriebspanels. FDZ Datenreport 2/2006.
- Bender, S./Hilzendegen, J./Rohwer, G./Rudolph, H. (1996): Die IAB-Beschäftigtenstichprobe 1975-1990. Beiträge zur Arbeitsmarkt- und Berufsforschung 197, Nürnberg.
- Bellmann, L. (2002): Das IAB-Betriebspanel. Konzeption und Anwendungsbereiche. In: Allgemeines Statistisches Archiv 86(2), 177-188.
- Brixy, U./Fritsch, M. (2002): Die Betriebsdatei der Beschäftigtenstatistik der Bundesanstalt für Arbeit. In: M. Fritsch & R. Grotz (Hrsg.), Das Gründungsgeschehen in Deutschland. Darstellung und Vergleich der Datenquellen, Heidelberg: Physica-Verlag, 55-78.
- Bundesagentur für Arbeit (2005): Statistik der sozialversicherungspflichtigen Beschäftigung und geringfügig entlohnerten Beschäftigung.
<http://www.pub.arbeitsamt.de/hst/services/statistik/000200/html/qualitaetsberichte/qualitaetsbericht-bst.pdf> (abgerufen am 07.08.2007).
- Bundesanstalt für Arbeit (1973) (Hrsg.): Verzeichnis der Wirtschaftszweige für die Statistik der Bundesanstalt für Arbeit, Nürnberg.
- Buttler, G./Fickel, N. (2002): Statistik mit Stichproben, Reinbek bei Hamburg: Rowohlt.
- Fischer, G./Wahse, J./Dahms, V./Frei, M./Riedmann, A./Janik, F. (2007): Standortbedingungen und Beschäftigung in den Regionen West- und Ostdeutschlands Ergebnisse des IAB-Betriebspanels 2006. IAB Forschungsbericht 5/2007.
- Gerlach, K./Hübler, O./Meyer, W. (1998) (Hrsg.): Ökonomische Analysen betrieblicher Strukturen und Entwicklungen. Das Hannoveraner Firmenpanel, Frankfurt/New York: Campus.
- Green, W. H. (2003): Econometric Analysis. Prentice Hall.
- Kölling, Arnd (2000): The IAB-Establishment Panel. In: Schmollers Jahrbuch 120, 291-300.
- Städele, D./Müller, D. (2006): Codebuch zum IAB-Betriebspanel, Version 1: Längsschnitt 1993 - 2005 (6. überarbeitete Auflage). FDZ Datenreport 6/2006.
- Statistisches Bundesamt (2003) (Hrsg.): Klassifikation der Wirtschaftszweige, Wiesbaden.
-

Anhang

Berechnung der Längsschnitthochrechnungsfaktoren

Das nachstehend in den Grundzügen beschriebene Verfahren wird für jeden einzelnen der Längsschnitte angewendet. Modifikationen ergeben sich hinsichtlich des Differenzierungsgrades der einzelnen Parameter in Anhängigkeit von den zur Verfügung stehenden Fallzahlen und in Anhängigkeit von dem abzubildenden Gesamtzeitraum.

Im ersten Schritt zur Berechnung der Längsschnitthochrechnungsfaktoren werden die am aktuellen Rand zur Grundgesamtheit des Querschnitts gehörenden Fälle an die Struktur der Grundgesamtheit (Anzahl der Betriebe und sozialversicherungspflichtig Beschäftigten am 30.06. des Vorjahres) angepasst. Eingangsfaktor ist dabei der Panelhochrechnungsfaktor des Vorjahres bzw. – bei neuen Betriebsnummern – der Querschnitthochrechnungsfaktor des laufenden Jahres. Dieser Schritt entspricht der in Kapitel 4.2 beschriebenen Querschnitthochrechnung, allerdings auf der Basis einer kleineren Fallzahl, da nur die Fälle einbezogen werden, die zum jeweiligen Längsschnitt und zum entsprechenden Querschnitt gehören. Als Basis für diesen Schritt dient die in der jeweiligen Querschnitthochrechnung verwendete Schichtungsmatrix.

Als nächstes werden die Eckzahlen für den Bestand an Betrieben für jedes einzelne vom Längsschnitt eingeschlossene Jahr sowie für die zwischen den Jahren „erloschenen“ und „neuen“ Betriebsnummern angepasst. Zum Bestand jedes Jahres gehören die Betriebe mit mindestens einem sozialversicherungspflichtigen Beschäftigten zum Stichtag. Als „neue“ Betriebsnummern gelten diejenigen aus dem Bestand, die zum Stichtag ein Jahr vorher keine sozialversicherungspflichtigen Beschäftigten hatten. Zu den „erloschenen“ Betriebsnummern zählen diejenigen aus dem Bestand, die zum Stichtag ein Jahr später keine sozialversicherungspflichtigen Beschäftigten mehr haben³⁴. Mit diesem Schritt wird das „Aufleben“ und „Absterben“ von Betriebsnummern im Zeitverlauf nachgebildet.

In einem dritten Schritt werden die am aktuellen Rand „überlebenden“ und die „erloschenen“ Betriebe von den Startwellen des jeweiligen Längsschnitts bis zur aktuellen Welle angepasst und für jedes Jahr des Längsschnittes die überlebenden und erloschenen Neugründungen kontrolliert.

³⁴ Dabei ist zu beachten, dass erloschene Betriebe laut Wellenkennzeichen, soweit dies im Rahmen der Kontaktaufnahme zum Interview erfahrbar ist, auch tatsächlich geschlossen wurden. Hingegen können sich hinter „erloschenen“ Betriebsnummern auch solche verbergen, die zum Stichtag keine sozialversicherungspflichtig Beschäftigte haben oder deren Betriebsnummer aus anderen Gründen nicht (mehr) verwendet wird (vgl. Kapitel 2.1).

Im vierten und fünften Schritt wird die Betriebsdynamik (Wachsen und Schrumpfen von Betrieben) berücksichtigt. Ein Wechsel der Betriebsgrößenklasse wird allerdings nur zwischen der Startwelle des jeweiligen Längsschnitts (bzw. der erstmaligen Befragung bei neuen Betriebsnummern) und dem jeweiligen Endjahr des Panelzeitraums berücksichtigt. Dazwischen liegende Veränderungen werden bei der Panelhochrechnung nicht kontrolliert. Wegen extrem kleiner Besetzungszahlen bestimmter Kombinationen werden darüber hinaus die theoretisch möglichen Kombinationen wie folgt zusammengefasst: Betrieb ist am aktuellen Rand in derselben Betriebsgrößenklasse wie zu Beginn bzw. ist gewachsen bzw. ist geschrumpft. Mit diesem Verfahren können die Entwicklungspfade von Betrieben daher nur näherungsweise abgebildet werden. Im vierten Schritt wird also die Betriebsdynamik für die Betriebe aus dem Bestand der Startwelle des jeweiligen Längsschnitts berücksichtigt. Im fünften Schritt wird dies für die während des Längsschnitts hinzu gekommenen „neuen“ Betriebsnummern gemacht.

Anschließend erfolgt die Korrektur eventueller disproportionaler Ausfälle in Abhängigkeit von Antworten auf einzelne Fragen aus der Erhebung des Vorjahres. Dazu werden mit Hilfe multivariater Schätzungen Auswahlwahrscheinlichkeiten in Abhängigkeit von den Fragebogenangaben der Vorwelle berechnet. Diese multivariaten Ausfallanalysen werden für jeden einzelnen Längsschnitt vorgenommen. Die Analysen verlaufen – mit Ausnahme des Längsschnitts ab 1993 und des Längsschnitts für das Verarbeitende Gewerbe in Ostdeutschland – getrennt für Ost und West.

In den Schritten sieben und acht wird eine Anpassung der zum jeweiligen Querschnitt gehörenden Fälle an die Vorgaben der Schichtungsmatrix für jedes einzelne vom Längsschnitt eingeschlossene Erhebungsjahr vorgenommen. Um die Zahl der Hochrechnungszellen zu begrenzen, werden nur die Betriebsgrößenklassen (siebter Schritt) sowie eine vereinfachte Branchenstruktur mit sieben Kategorien (achter Schritt) berücksichtigt.

Die acht Gewichtungsschritte werden iterativ wiederholt.

Die Berücksichtigung der oben genannten Parameter bei der Panelhochrechnung kann in Einzelfällen zu extrem hohen Gewichtungsfaktoren führen. Um die Ausreißerproblematik zu begrenzen werden für die Panelhochrechnung nur Gewichtungsfaktoren zwischen eins³⁵ und max. 4.000 zugelassen. Dies kann zur Folge haben, dass für einzelne Merkmalsausprägungen die theoretisch erforderlichen Soll-Zahlen durch die Gewichtung nicht erreicht werden.

³⁵ Bis 2001 wurden auch Längsschnittshochrechnungsfaktoren kleiner eins zugelassen (zu den Ursachen vgl. auch Kapitel 4.2).

Tabelle A: Auswahlwahrscheinlichkeiten nach Betriebsgröße

Anzahl der sozialversicherungs- pflichtig Beschäftigten	durchschnittliche Auswahlwahrscheinlichkeit
1 bis 4	0.0011
5 bis 9	0.0015
10 bis 19	0.0030
20 bis 49	0.0089
50 bis 99	0.0153
100 bis 199	0.0304
200 bis 499	0.0862
500 bis 999	0.1504
1000 bis 4999	0.8765
5000 u. m.	0.9127
Insgesamt	0.0043

Quelle: IAB-Betriebspanel 1993

Tabelle B: Branchengliederung nach 16 Wirtschaftszweigen für die Stichprobenziehung und Hochrechnung von Welle 1993 bis 1999

Code	Branche	Code laut Fragebogen 1993-1998	Code laut Fragebogen 1999
1	Land-, Forstwirtschaft	1	010
2	Energie, Bergbau	2	020
3	Grundstoffverarbeitung	3-6	030-060
4	Investitionsgüter	7-12	071, 072, 073, 080, 091, 092, 100, 110, 120
5	Verbrauchsgüter	13-16	130-160
6	Baugewerbe	17-18	170-180
7	Handel	19	190
8	Verkehr, Nachrichten	20	200
9	Kreditgewerbe	21	210
10	Versicherungen	22	220
11	Gaststätten, Heime	23-25	230-250
12	Bildung, Verlage	26-27	260-270
13	Gesundheitswesen	28	280
14	Freie Berufe u.ä.	29-34	290-340
15	sonstige Dienstleistungen	35	350
16	Organisationen ohne Erwerbscharakter/ Staat	36-41	360-410

In der Welle 1999 wird die 41 Branchenklassifizierung nicht auf 2-Steller-Ebene, sondern differenzierter auf 3-Steller-Ebene abgefragt. Dies war notwendig, um Betriebe des Verarbeitenden Gewerbes für die Aufstockung des IW Halle identifizieren zu können. Die stärkste Differenzierung betraf die Fragebogen-codes 7 und 9.

Tabelle C: Branchengliederung nach 20 Wirtschaftszweigen für die Stichprobenziehung und Hochrechnung von Welle 2000 bis 2003

Code	Branche	Code WZ93	Code laut Fragebogen 2000-2003
1	Land- und Forstwirtschaft, Fischerei und Fischzucht	01, 02, 05	01
2	Bergbau und Gewinnung von Steinen und Erden, Energie und Wasserversorgung	10-14, 40, 41	02
3	Nahrungs- und Genussmittel	15-16	03
4	Verbrauchsgüter	17-19, 21-22, 36	04-05, 18
5	Produktionsgüter	20, 23-27, 37	06-11
6	Investitions- und Gebrauchsgüter	28-35	12-17
7	Baugewerbe	45	19, 20
8	Handel, Instandhaltung und Reparatur von Kraftfahrzeugen und Gebrauchsgütern	50-52	21-23
9	Verkehr und Nachrichtenübermittlungen	60-64	24, 25
10	Kredit- und Versicherungsgewerbe	65-67	26, 27
11	Gastgewerbe	55	33
12	Erziehung und Unterricht	80	34
13	Gesundheits- Veterinär- und Sozialwesen	85	35
14	Datenverarbeitung und Datenbanken	72	28
15	Forschung und Entwicklung	73	29
16	Rechts- und Steuer- und Unternehmensberatung, Markt- und Meinungsforschung, Beteiligungsgesellschaften, Werbung	741, 744	30
17	Grundstücks- und Wohnungswesen	70	31
18	Vermietung beweglicher Sachen, sonstige Dienstleistungen überwiegend für Unternehmen	71, 742-743, 745-748	32
19	sonstige Dienstleistungen	90, 92-93	36-38
20	Öffentliche Verwaltung, Verteidigung, Sozialversicherung, private Haushalte, extraterritoriale Organisationen und Körperschaften, Verbände, Interessenvertretungen	75, 91, 95, 99	39-41

Tabelle D: Branchengliederung nach 17 Wirtschaftszweigen für die Stichprobenziehung und Hochrechnung ab der Welle 2004

Code	Branche	Code WZ2003	Code laut Fragebogen ab 2004
1	Land- und Forstwirtschaft, Fischerei und Fischzucht	01, 02, 05	01
2	Bergbau und Gewinnung von Steinen und Erden, Energie und Wasserversorgung	10-14, 40, 41	02
3	Nahrungs- und Genussmittel	15-16	03
4	Verbrauchsgüter (ohne Holzgewerbe)	17-19, 21-22, 36	04-05, 18
5	Produktionsgüter (einschl. Holzgewerbe)	20, 23-27, 37	06-11
6	Investitions- und Gebrauchsgüter	28-35	12-17
7	Baugewerbe	45	19, 20
8	Handel, Instandhaltung und Reparatur von Kraftfahrzeugen und Gebrauchsgütern	50-52	21-23
9	Verkehr und Nachrichtenübermittlungen	60-64	24, 25
10	Kredit- und Versicherungsgewerbe	65-67	26, 27
11	Gastgewerbe	55	33
12	Erziehung und Unterricht	80	34
13	Gesundheits- Veterinär- und Sozialwesen	85	35
14	Unternehmensnahe Dienstleistungen	70-74	28-32
15	Sonstige Dienstleistungen	90, 92-93	36-38
16	Organisationen ohne Erwerbszweck	91	39
17	Öffentliche Verwaltung	75	41

Tabelle E: Regionale Aufstockungen 1993-2007

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Schleswig-Holstein										X	X	X	X	X	X
Hamburg								X	X						
Niedersachsen								X	X	X	X	X	X	X	X
Bremen								X	X	X	X	X	X	X	X
Nordrhein-Westfalen								X	X	X	X	X	X	X	X
Hessen									X	X	X	X	X	X	X
Rheinland-Pfalz								X	X	X	X	X	X	X	X
Baden-Württemberg								X	X	X	X	X	X	X	X
Bayern								X	X	X	X	X	X	X	X
Saarland									X	X	X	X	X	X	X
Berlin				X	X	X	X	X	X	X	X	X	X	X	X
Brandenburg				X	X	X	X	X	X	X	X	X	X	X	X
Mecklenburg-Vorpommern				X	X	X	X	X	X	X	X	X	X	X	X
Sachsen				X	X	X	X	X	X	X	X	X	X	X	X
Sachsen-Anhalt								X	X	X	X	X	X	X	X
Thüringen				X	X	X	X	X	X	X	X	X	X	X	X
Verarbeitendes Gewerbe Ost							X	X	X	X	X	X	X	X	X

Tabelle F: Panelfälle

	_94	_95	_96	_97	_98	_99	_00	_01	_02	_03	_04	_05	_06
pan93...	3739	3614	8004	3252	3221	2872	2957	3243	3053	3230	3501	3647	3871
pan96...				8048	7537	6718	6581	6842	6441	6705	7086	7252	7562
pan00...								12135	10528	10006	9987	9773	9772
pan03...											14179	13154	12448

Tabelle G: Panelfälle mit gültigen Interviews in jedem Jahr

	_94	_95	_96	_97	_98	_99	_00	_01	_02	_03	_04	_05	_06
pan93...	3489	2916	2381	1989	1701	1412	1222	1053	892	762	656	580	509
pan96...				6658	5483	4565	3943	3427	2952	2576	2257	1987	1740
pan00...								10840	8746	7281	6347	6636	4841
pan03...											12769	10761	9108

Tabelle H: Fehlertoleranztabelle

Fehlertoleranztabelle

p = (%)	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95
n = 100	(-)	(-)	(-)	(-)	12.2	13.0	13.5	13.9	14.1	14.1	14.1	13.9	13.5	13.0	12.2	11.3	10.1	8.5	6.2
200	(-)	(-)	7.1	8.0	8.7	9.2	9.5	9.8	9.9	10.0	9.9	9.8	9.5	9.2	8.7	8.0	7.1	6.0	4.4
300	(-)	4.9	5.8	6.5	7.1	7.5	7.8	8.0	8.1	8.2	8.1	8.0	7.8	7.5	7.1	6.5	5.8	4.9	3.6
400	(-)	4.2	5.0	5.7	6.1	6.5	6.7	6.9	7.0	7.1	7.0	6.9	6.7	6.5	6.1	5.7	5.0	4.2	3.1
500	(-)	3.8	4.5	5.1	5.5	5.8	6.0	6.2	6.3	6.3	6.3	6.2	6.0	5.8	5.5	5.1	4.5	3.8	2.8
600	(-)	3.5	4.1	4.6	5.0	5.3	5.5	5.7	5.7	5.8	5.7	5.7	5.5	5.3	5.0	4.6	4.1	3.5	2.5
700	2.3	3.2	3.8	4.3	4.6	4.9	5.1	5.2	5.3	5.3	5.2	5.1	4.9	4.6	4.3	3.8	3.2	2.3	
800	2.2	3.0	3.6	4.0	4.3	4.6	4.8	4.9	5.0	5.0	4.9	4.8	4.6	4.3	4.0	3.6	3.0	2.2	
900	2.1	2.8	3.4	3.8	4.1	4.3	4.5	4.6	4.7	4.7	4.6	4.5	4.3	4.1	3.8	3.4	2.8	2.1	
1.000	1.9	2.7	3.2	3.6	3.9	4.1	4.3	4.4	4.4	4.4	4.4	4.3	4.1	3.9	3.6	3.2	2.7	1.9	
1.100	1.9	2.6	3.0	3.4	3.7	3.9	4.1	4.2	4.2	4.3	4.2	4.2	4.1	3.9	3.7	3.4	3.0	2.6	1.9
1.200	1.8	2.4	2.9	3.3	3.5	3.7	3.9	4.0	4.1	4.1	4.1	4.0	3.9	3.7	3.5	3.3	2.9	2.4	1.8
1.300	1.7	2.4	2.8	3.1	3.4	3.6	3.7	3.8	3.9	3.9	3.9	3.8	3.7	3.6	3.4	3.1	2.8	2.4	1.7
1.400	1.6	2.3	2.7	3.0	3.3	3.5	3.6	3.7	3.8	3.8	3.8	3.7	3.6	3.5	3.3	3.0	2.7	2.3	1.6
1.500	1.6	2.2	2.6	2.9	3.2	3.3	3.5	3.6	3.6	3.7	3.6	3.6	3.5	3.3	3.2	2.9	2.6	2.2	1.6
1.600	1.5	2.1	2.5	2.8	3.1	3.2	3.4	3.5	3.5	3.5	3.5	3.4	3.2	3.1	2.8	2.5	2.1	1.5	
1.700	1.5	2.1	2.4	2.7	3.0	3.1	3.3	3.4	3.4	3.4	3.4	3.3	3.1	3.0	2.7	2.4	2.1	1.5	
1.800	1.5	2.0	2.4	2.7	2.9	3.1	3.2	3.3	3.3	3.3	3.3	3.2	3.1	2.9	2.7	2.4	2.0	1.5	
1.900	1.4	1.9	2.3	2.6	2.8	3.0	3.1	3.2	3.2	3.2	3.2	3.1	3.0	2.8	2.6	2.3	1.9	1.4	
2.000	1.4	1.9	2.3	2.5	2.7	2.9	3.0	3.1	3.1	3.2	3.1	3.1	3.0	2.9	2.7	2.5	2.3	1.9	1.4
2.500	1.2	1.7	2.0	2.3	2.4	2.6	2.7	2.8	2.8	2.8	2.8	2.7	2.6	2.4	2.3	2.0	1.7	1.2	
3.000	1.1	1.5	1.8	2.1	2.2	2.4	2.5	2.5	2.6	2.6	2.6	2.5	2.5	2.4	2.2	2.1	1.8	1.1	
4.000	1.0	1.3	1.6	1.8	1.9	2.0	2.1	2.2	2.2	2.2	2.2	2.2	2.1	2.0	1.9	1.8	1.6	1.3	1.0
6.000	0.8	1.1	1.3	1.5	1.6	1.7	1.7	1.8	1.8	1.8	1.8	1.7	1.7	1.6	1.5	1.3	1.1	0.8	
8.000	0.7	0.9	1.1	1.3	1.4	1.4	1.5	1.5	1.6	1.6	1.6	1.5	1.5	1.4	1.4	1.3	1.1	0.9	0.7
10.000	0.6	0.8	1.0	1.1	1.2	1.3	1.3	1.4	1.4	1.4	1.4	1.3	1.3	1.2	1.1	1.0	0.8	0.6	
15.000	0.5	0.7	0.8	0.9	1.0	1.1	1.1	1.1	1.1	1.2	1.1	1.1	1.1	1.0	0.9	0.8	0.7	0.5	
20.000	0.4	0.6	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9	0.9	0.8	0.7	0.6	0.4
25.000	0.4	0.5	0.6	0.7	0.8	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.8	0.8	0.7	0.6	0.5	0.4

Beispiel: Eine Stichprobe vom Umfang $n = 2.000$ lieferte einen Anteil von 30 % Bausparern. Aus der Tabelle entnimmt man die Fehlergrenzen ± 2.9 % (markierter Wert). Mit einer Wahrscheinlichkeit von 95.5 % liegt dann der wahre Anteil in der Grundgesamtheit bei $30 \% \pm 2.9$ %, d. h. zwischen 27.1 % und 32.9 %.

Die niedrigen Anteilswerte aus kleinen Stichproben, bei denen keine Fehlertoleranz angegeben ist, sind nur eingeschränkt interpretierbar, weil diese Fehlertoleranzen mehr als die Hälfte des Anteilswertes betragen (z.B. ± 8.5 % für $n = 100$ und $p = 10$ %).

$$p - t\sigma \leq p \leq p + t\sigma$$

$$\sigma = \sqrt{2} \sqrt{\frac{p(100-p)}{n}}$$

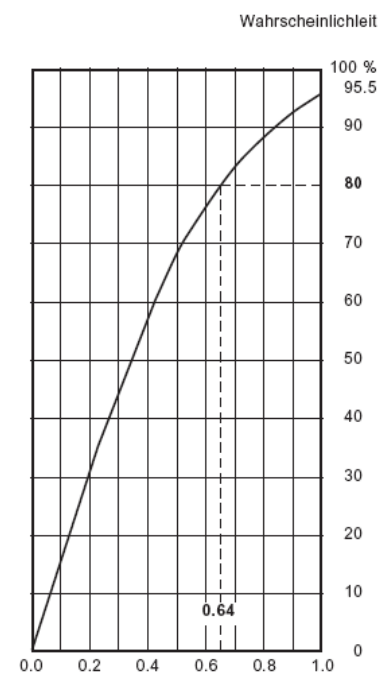
$t = 2 \Rightarrow$ Sicherheitsgrad 95.5 %
 P = Merkmalsanteil in Grundgesamtheit

$$\sqrt{2} = \text{Design-Faktor}$$

p = Merkmalsanteil in der Stichprobe (in %)
 n = Stichprobenumfang
 Bitte beachten: Bei Teilgruppenauswertungen muß in der Regel auf den gesamten Stichprobenumfang prozentuiert werden!



Kurve zur Bestimmung der Fehlertoleranzen bei reduziertem Sicherheitsniveau.



Reduktionsverfahren zu den in der Tabelle ausgewiesenen Fehlergrenzen.

Beispiel:
 Einem Sicherheitsniveau von 80 % entspricht in der Grafik der Reduktionsfaktor 0.64 (vgl. gestrichelte Linie).
 Das nebenstehende Beispiel Bausparer lautet dann: Mit einer Wahrscheinlichkeit von 80 % liegt der wahre Anteil bei $30 \% \pm 1.9$ % (2.9 % mal $0.64 = 1.9$ %), d. h. zwischen 28.1 % und 31.9 %.

Imprint

FDZ *Methodenreport*

No. 01/2008

Publisher

The Research Data Centre (FDZ)
of the Federal Employment Service
in the Institute for Employment Research
Regensburger Str. 104
D-90478 Nuremberg

Editorial staff

Stefan Bender, Dagmar Herrlinger

Technical production

Dagmar Herrlinger

Copyright

Reproduction – also in parts – only with permission of
the FDZ

Download

http://doku.iab.de/fdz/reporte/2008/MR_01-08.pdf

Internet

<http://fdz.iab.de/>

Corresponding author

Dana Müller, Institute for Employment Research,
Regensburger Str. 104, D-90478 Nuremberg
Phone: +49-(0)911/179-2409
E-Mail: dana.mueller@iab.de